

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:23:41 ; Search time 38.6861 Seconds
(without alignments)
287.115 Million cell updates/sec

Title: US-09-894-030-3
Perfect score: 545
Sequence: 1 MEKLFKEVKLELENQKSG.....DDFDLDVVKVQDSKITPQ 100

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	304	55.8	56	AAV03205	Amino acid sequenc
2	211	38.7	37	AAV03207	Amino acid sequenc
3	202	37.1	37	AAV03218	Amino acid sequenc
4	142.5	26.1	576	AA12229	TrpE/androgen rece
5	142.5	26.1	918	AA12223	Human androgen rec
6	142.5	26.1	918	AAV33491	Human androgen rec
7	142.5	26.1	919	AA193109	Human androgen rec
8	142.5	26.1	919	AA190996	Human androgen rec
9	142.5	26.1	919	AAW14783	Androgen receptor.
10	142.5	26.1	919	AAV78914	Human androgen rec
11	140	25.7	94	AA171269	Human gene 1-encod

12	135.5	24.9	240	16	AA178738	Murine mv1 protein
13	135.5	24.9	240	18	AAW26594	Murine BMP-12 homo
14	135.5	24.9	240	22	AAE10984	Murine partial mv1
15	134.5	24.7	101	22	ABB27649	Human peptide #300
16	134.5	24.7	101	22	ABB32819	Peptide #325 encod
17	134.5	24.7	101	22	ABB18301	Protein #300 encod
18	134.5	24.7	101	22	AAW53626	Human brain expres
19	134.5	24.7	101	22	AAW66005	Human bone marrow
20	134.5	24.7	101	22	AAW13874	Peptide #308 encod
21	134.5	24.7	101	22	AAW26281	Peptide #318 encod
22	134.5	24.7	101	22	AAW01617	Peptide #299 encod
23	129.5	23.8	738	19	AAW56163	New DNA sequence i
24	129	23.7	268	18	AAW41640	Human calpain smal
25	129	23.7	268	22	AAW86127	Human calpain 30kD
26	127.5	23.4	118	22	ABB58597	Drosophila melanog
27	127	23.3	399	22	ABB60010	Drosophila melanog
28	126.5	23.2	348	22	ABB59015	Drosophila thalia
29	126	23.1	258	21	AAG51723	Arabidopsis thalia
30	126	23.1	307	18	AAW18563	Novel fusion prote
31	125	22.9	50	18	AAW13633	Apolipoprotein (a)
32	125	22.9	161	16	AAW65182	GDF-7 C-terminal r
33	125	22.9	569	22	AAW79339	Human protein SEQ
34	125	22.9	569	22	AAW79340	Human protein SEQ
35	124.5	22.8	321	15	AAW54683	Human fibrillar in.
36	124.5	22.8	334	21	AAW43918	Human cancer assoc
37	124.5	22.8	357	22	ABB29912	Peptide #2563 enco
38	124.5	22.8	357	22	ABB35090	Peptide #2596 enco
39	124.5	22.8	357	22	ABB20509	Protein #2508 enco
40	124.5	22.8	357	22	AAW55912	Human brain expres
41	124.5	22.8	357	22	AAW68282	Human bone marrow
42	124.5	22.8	357	22	AAW16105	Peptide #2539 enco
43	124.5	22.8	357	22	AAW28596	Peptide #2633 enco
44	124.5	22.8	357	22	AAW03832	Peptide #2514 enco
45	124	22.8	50	18	AAW13632	Apolipoprotein (a)

ALIGNMENTS

RESULT 1	
AAV03205	AAV03205 standard; Protein; 56 AA.
ID	AAV03205 standard; Protein; 56 AA.
XX	
AC	AAV03205;
XX	
DT	03-AUG-1999 (first entry)
XX	
DE	Amino acid sequence of sunA protein.
XX	
KW	Sublancin 168; antimicrobial; food preservative; pro-sublancin 168;
KW	Gram-positive bacteria; pre-sublancin 168; sunA; sunt.
XX	
OS	Bacillus subtilis.
XX	
PN	WO9903352-A1.
XX	
PD	28-JAN-1999.
XX	
PF	17-JUL-1998; 98WO-US14547.
XX	
PR	18-JUL-1997; 97US-0053035.
XX	
PA	(UYMA-) UNIV MARYLAND BALTIMORE.
XX	
PI	Hansen JN;
XX	
DR	WPI; 1999-131752/11.
DR	N-PSDB; AAX28631.
XX	
PT	New antimicrobial peptide, sublancin 168, from Bacillus subtilis -
PT	used for, e.g. treatment of infections caused by Gram negative
PT	bacteria and as food preservative
XX	

XX AAR12229;
AC 20-AUG-1991 (first entry)
DE TrpE/androgen receptor N-terminal domain fusion protein.
XX androgen receptor; AR; DNA-binding protein; steroid hormone.
KW Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT Protein 1..323
FT Region /label= 33kD trpE protein
FT /label= 324..329
FT Protein /label= 6 amino acid linker
FT 330..571
FT Region /label= N-terminal region of AR
FT 572..576
FT /label= 5 amino acid linker
XX
PN WO9107423-A.
XX
PD 30-MAY-1991.
XX
PF 19-OCT-1990; 90WO-US06015.
XX
PR 17-NOV-1989; 89US-0438775.
XX
XX (ARCH-) ARCH DEV CORP.
PA Liao S, Chang C;
PI WPI; 1991-178048/24.
XX N-PSDB; AAQ12007.
XX Androgen receptor and TR2 DNA binding proteins - DNA sequences
PT and antibodies for detection and quantification methods
XX
PS Example 13; Fig 9; 79pp; English.
XX
XX To express an androgen receptor fusion protein in E.coli, the PATH
CC expression system was used. The trpE is insoluble so partially
CC purified induced fusion protein is obtained by simply lysing the
CC E.coli and precipitating the insoluble fusion protein. The fusion
CC protein was used for immunisation to obtain monoclonal anti-AR
CC antibodies.
XX
SQ Sequence 576 AA;
Query Match 26.1%; Score 142.5; DB 12; Length 576;
Best Local Similarity 50.8%; Pred. No. 9.9e-07;
Matches 32; Conservative 3; Mismatches 13; Indels 15; Gaps 3;
QY 18 GSGLGKQAQCAALW--LQCASGGTI--GCGGAVACQNYRQFCRGGGGGGGGGGGGG 73
DB 416 GSGSPSAASSSWHTLFTAEQGQLYGPCGGG-----GGGGGGGGGGGGGGG 464
QY 74 GGG 76
DB 465 GGG 467
RESULT 5
AAR12223
ID AAR12223 standard; Protein; 918 AA.
XX
AC AAR12223;
XX
DT 20-AUG-1991 (first entry)
XX
DE Human androgen receptor.
XX

KW hAR; DNA-binding protein; steroid hormone.
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT Domain 556..626
FT /label= DNA-binding domain
FT /note= "cysteine-rich"
XX
PN WO9107423-A.
XX
PD 30-MAY-1991.
XX
PF 19-OCT-1990; 90WO-US06015.
XX
PR 17-NOV-1989; 89US-0438775.
XX
XX (ARCH-) ARCH DEV CORP.
PA Liao S, Chang C;
PI WPI; 1991-178048/24.
XX N-PSDB; AAQ12001.
XX Androgen receptor and TR2 DNA binding proteins - DNA sequences
PT and antibodies for detection and quantification methods
XX
PS Claim 25; Fig 3; 79pp; English.
XX
XX This sequence was deduced from a cDNA clone isolated by screening
CC commercially available human testis and prostate lambda gt11 cDNA
CC libraries. The sequence is very similar to that of rat AR and in
CC the DNA-binding domain it is identical to that of rat AR DNA-binding
CC domain. Homology comparisons with other known steroid receptors
CC indicate that hAR is more closely related to glucocorticoid,
CC mineralo-corticoid and progesterone receptors than to v-erb-A or to
CC receptors for oestrogen, vitamin D and thyroid hormones.
XX
SQ Sequence 918 AA;
Query Match 26.1%; Score 142.5; DB 12; Length 918;
Best Local Similarity 50.8%; Pred. No. 1.6e-06;
Matches 32; Conservative 3; Mismatches 13; Indels 15; Gaps 3;
QY 18 GSGLGKQAQCAALW--LQCASGGTI--GCGGAVACQNYRQFCRGGGGGGGGGGGGG 73
DB 417 GSGSPSAASSSWHTLFTAEQGQLYGPCGGG-----GGGGGGGGGGGGGGG 465
QY 74 GGG 76
DB 466 GGG 468
RESULT 6
AAY33491
ID AAY33491 standard; Protein; 918 AA.
XX
AC AAY33491;
XX
DT 19-JAN-2000 (first entry)
XX
DE Human androgen receptor protein.
XX
XX Proapoptotic; dependence domain; p75NTR; androgen receptor; DCC;
KW huntingtin polypeptide; Machado-Joseph disease; SCA1; SCA2; SCA6;
KW atrophin-1; cell death; apoptosis; Huntington's disease; head trauma;
KW Alzheimer's disease; Kennedy's disease; spinocerebellar ataxia; stroke;
KW dentatorubropallidolusian atrophy; cell proliferation; cell survival;
KW neoplastic; malignant; autoimmune; fibrotic.
XX
OS Homo sapiens.
XX
PN WO9945944-A1.

Db 386 KLENPLDYGSAWAAAAA---QCRYGDLASLHGAGAGCGSPSAAASSSWHTLFTAEEG 442
Qy 55 -----CRGGGGGGGGGGGGGGGGGGGG 76
Db 443 QLYGPGGGGGGGGGGGGGGGGGGGGG 469

RESULT 11
AAG711269
ID AAG711269 standard; Protein; 94 AA.
XX
AC AAG711269;
XX
DT 30-JUL-2001 (first entry)
XX
DE Human gene 1-encoded secreted protein HTEGI42, SEQ ID NO:118.
XX
KW Human; secreted protein; proliferative disorder; cancer; chromosome 15;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; tumour;
KW endocrine disorder; infection; wound healing; vulneryary;
KW cell culture; chemotaxis; food additive;
KW binding partner identification.
XX
OS Homo sapiens.
XX
PN WO200132674-A1.
XX
PD 10-MAY-2001.
XX
PF 25-OCT-2000; 2000WO-US29360.
XX
PR 29-OCT-1999; 99US-0162211.
PR 30-JUN-2000; 2000US-0215138.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Young PE, Moore PA;
XX
DR WPI; 2001-291051/30.
DR N-PSDB; AAH31376.
XX
PT New nucleic acid molecule encoding a human secreted protein, useful for
PT preventing, treating or ameliorating medical conditions such as
PT rheumatoid arthritis, Alzheimer's disease and microbial infections -
XX
PS Claim 11; Page 496; 581pp; English.
XX
AAH31349-AAH31428 represent cDNAs corresponding to 26 human secreted
protein genes, and AAG71243-AAG71319 represent the proteins they encode.
AAG71320-AAG71403 represent human secreted protein fragments. The genes
and their corresponding secreted proteins are useful for preventing,
treating or ameliorating medical conditions, e.g., by protein or gene
therapy. Pathological conditions can be diagnosed by determining the
amount of the new protein in a sample or by determining the presence of
mutations in the new genes. Specific uses are described for each of the
52 genes, based on the tissues in which they are most highly expressed,
and include developing products for the diagnosis or treatment of
proliferative disorders, cancer, tumours, foetal and developmental
abnormalities, haematopoietic disorders, diseases of the immune system,
AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
allergies, neurological disorders (e.g., Alzheimer's disease,
Parkinson's disease), cognitive disorders, schizophrenia, asthma,
skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
cardiovascular disorders, angiogenic disorders, kidney disorders,
gastrointestinal disorders, pregnancy-related disorders, endocrine
disorders, and infections. The proteins can also be used to aid wound

CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein of the invention.
XX
SQ Sequence 94 AA;

Query Match 25.7%; Score 140; DB 22; Length 94;
Best Local Similarity 59.6%; Pred. NO. 2.7e-07;
Matches 28; Conservative 1; Mismatches 4; Indels 14; Gaps 2;

QY 33 CASGGTIGCGGAVACQNYRQFCRGGGGGGGGGGGGGGGGGGGMSK 79
Db 23 CQGGG--GGGGG-----GGGGGGGGGGGGGGGGGGGVDK 55

RESULT 12

AAR78738
ID AAR78738 standard; Protein; 240 AA.

XX AAR78738;

DT 23-NOV-1995 (first entry)

DE Murine mV1 protein.

OS Bone morphogenetic protein; mV1; tendon; ligament.

XX Mus musculus.

Key Location/Qualifiers
Misc-difference 7 /label= V,A,E,G

Misc-difference 8 /label= S,P T,A

Misc-difference 15 /label= S,R

Misc-difference 16 /label= L,P,Q,R

Misc-difference 26 /label= C,W

Misc-difference 37 /label= V,A,D,G

Misc-difference 103 /label= V,A,E,G

Misc-difference 104 /label= Q,K,E

XX WO9516035-A.

PN 15-JUN-1995.

PD 06-DEC-1994; 94WO-US14030.

XX 02-NOV-1994; 94US-0333576.

PR 07-DEC-1993; 93US-0164103.

PR 25-MAR-1994; 94US-0217780.

XX (GEMY) GENETICS INST INC.

PA (HARD) HARVARD COLLEGE.

XX Celeste AJ, Melton DA, Rosen VA, Thomsen GH, Wolfman NM;

PI Wozney JM;

XX WPI; 1995-224320/29.

DR N-PSDB; AAQ96223.

XX

Thu Oct 24 14:39:59 2002

QY 66 GGGGGGGGGGMSK 79
| | | | | | | | | |
Db 48 GGGGGGGGGRRRK 61

Search completed: October 24, 2002, 09:27:58
Job time : 41.6861 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:28:06 ; Search time 152.555 Seconds
(without alignments)
230.724 Million cell updates/sec

Title: US-09-894-030-3
Perfect score: 545
Sequence: 1 MEKLFKEVKLELENQKSG.....DDFDLDVVKVSKQDSKITPQ 100

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main:*
1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US08_COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/1/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/1/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/1/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/1/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/1/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/1/paa/US100_COMB.pep:*
25: /cgn2_6/ptodata/1/paa/US101_COMB.pep:*
26: /cgn2_6/ptodata/1/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	545	100.0	100	22	US-09-893-499-3 Sequence 3, Appli
2	545	100.0	100	22	US-09-893-600-3 Sequence 3, Appli
3	545	100.0	100	22	US-09-894-030-3 Sequence 3, Appli
4	304	55.8	56	1	PCT-US98-14547-5 Sequence 5, Appli
5	304	55.8	56	1	PCT-US98-14547-5 Sequence 5, Appli
6	304	55.8	56	18	US-09-462-478A-5 Sequence 5, Appli
7	211	38.7	37	1	PCT-US98-14547-7 Sequence 7, Appli

8	211	38.7	37	1	PCT-US98-14547-7	Sequence 7, Appli
9	211	38.7	37	18	US-09-462-478A-7	Sequence 7, Appli
10	202	37.1	37	1	PCT-US98-14547-18	Sequence 18, Appli
11	202	37.1	37	1	PCT-US98-14547-18	Sequence 18, Appli
12	202	37.1	37	18	US-09-462-478A-18	Sequence 18, Appli
13	144	26.4	396	4	US-08-085-126-56	Sequence 56, Appli
14	144	26.4	396	8	US-08-438-114-56	Sequence 56, Appli
15	143	26.2	276	18	US-09-417-507-22220	Sequence 22220, A
16	142.5	26.1	918	1	PCT-US99-05250-11	Sequence 11, Appli
17	142.5	26.1	919	19	US-09-515-886-2	Sequence 2, Appli
18	142.5	26.1	919	19	US-09-538-092-895	Sequence 895, App
19	142.5	26.1	923	18	US-09-497-822-19	Sequence 19, Appli
20	142.5	26.1	923	18	US-09-497-822A-19	Sequence 19, Appli
21	141	25.9	175	18	US-09-417-507-43604	Sequence 43604, A
22	141	25.9	209	21	US-09-708-427-10485	Sequence 10485, A
23	141	25.9	221	21	US-09-708-427-10484	Sequence 10484, A
24	140	25.7	94	1	PCT-US00-29360-118	Sequence 118, App
25	140	25.7	94	1	PCT-US01-11988-197	Sequence 197, App
26	140	25.7	94	1	PCT-US02-08123-1836	Sequence 1836, Ap
27	140	25.7	94	1	PCT-US02-08277-1207	Sequence 1207, Ap
28	140	25.7	94	1	PCT-US02-08278-1752	Sequence 1752, Ap
29	140	25.7	94	22	US-09-833-245-197	Sequence 197, Appl
30	139.5	25.6	285	18	US-09-417-507-22221	Sequence 22221, A
31	138	25.3	188	18	US-09-417-507-22219	Sequence 22219, A
32	135.5	24.9	240	7	US-08-333-576A-30	Sequence 30, Appl
33	135.5	24.9	240	7	US-08-362-670-30	Sequence 30, Appl
34	134.5	24.7	101	1	PCT-US01-00663-26550	Sequence 26550, A
35	134.5	24.7	101	22	US-09-864-761-33599	Sequence 33599, A
36	134.5	24.7	101	26	US-60-236-359-14971	Sequence 14971, A
37	134.5	24.7	1084	17	US-09-394-272-9	Sequence 9, Appli
38	134	24.6	708	16	US-09-248-796-16456	Sequence 16456, A
39	134	24.6	708	26	US-60-096-409-16456	Sequence 16456, A
40	133.5	24.5	169	18	US-09-417-507-25745	Sequence 25745, A
41	133	24.4	252	18	US-09-417-507-43603	Sequence 43603, A
42	133	24.4	262	21	US-09-733-089-9222	Sequence 9222, Ap
43	133	24.4	262	22	US-09-816-660-9222	Sequence 9222, Ap
44	132.5	24.3	201	1	PCT-US01-00358-5	Sequence 5, Appli
45	132.5	24.3	201	1	PCT-US99-17885-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-893-499-3
; Sequence 3, Application US/09893499
; GENERAL INFORMATION:
; APPLICANT: Hansen, J. Norman
; TITLE OF INVENTION: Construction and Screening of Lantibody Display Libraries
; FILE REFERENCE: 108172-00056
; CURRENT APPLICATION NUMBER: US/09/893,499
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/215,449
; PRIOR FILING DATE: 2000-06-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: The peptide sequence of sunA-PG20-SL.
US-09-893-499-3

Query Match	100.0%	Score 545;	DB 22;	Length 100;
Best Local Similarity	100.0%	Pred. No. 1.4e-49;		
Matches 100;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MEKLFKEVKLELENQKSGGLGKAQCAALWLQACSGGTICGGGAVACQNYRQFCRGGG	60	
Db	1	MEKLFKEVKLELENQKSGGLGKAQCAALWLQACSGGTICGGGAVACQNYRQFCRGGG	60	
QY	61	GGGGGGGGGGGGGMSKDFDLDVVKVSKQDSKITPQ	100	

; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Bacillus subtilis 168
; FEATURE:
; OTHER INFORMATION: Pre-Sublancin 168
US-09-462-478A-5

Query Match 55.8%; Score 304; DB 18; Length 56;
Best Local Similarity 100.0%; Pred. No. 2.2e-24;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEKLFKEVKLEENQKSGGLGKAQCAALWLCASGGTIGCGGAVACQNYRQFCR 56
|||||
Db 1 MEKLFKEVKLEENQKSGGLGKAQCAALWLCASGGTIGCGGAVACQNYRQFCR 56
|||||

RESULT 7
PCT-US98-14547-7
; Sequence 7, Application PC/TUS9814547B
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND
; TITLE OF INVENTION: SUBLANCIN LANTIBIOTIC PRODUCED BY BACILLUS SUBTILIS 168
; FILE REFERENCE: 8172-8068
; CURRENT APPLICATION NUMBER: PCT/US98/14547B
; CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: US/60/053,035
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Bacillus subtilis 168
; FEATURE:
; OTHER INFORMATION: Prosublancin 168
PCT-US98-14547-7

Query Match 38.7%; Score 211; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 9.3e-15;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 GLGKAQCAALWLCASGGTIGCGGAVACQNYRQFCR 56
|||||
Db 1 GLGKAQCAALWLCASGGTIGCGGAVACQNYRQFCR 37
|||||

RESULT 8
PCT-US98-14547-7
; Sequence 7, Application PC/TUS9814547C
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND
; TITLE OF INVENTION: SUBLANCIN LANTIBIOTIC PRODUCED BY BACILLUS SUBTILIS 168
; FILE REFERENCE: 8172-8072
; CURRENT APPLICATION NUMBER: PCT/US98/14547C
; CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: US/60/053,035
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Bacillus subtilis 168
; FEATURE:
; OTHER INFORMATION: Prosublancin 168
PCT-US98-14547-7

Query Match 38.7%; Score 211; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 9.3e-15;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 GLGKAQCAALWLCASGGTIGCGGAVACQNYRQFCR 56
|||||
Db 1 GLGKAQCAALWLCASGGTIGCGGAVACQNYRQFCR 37
|||||

RESULT 9
US-09-462-478A-7
; Sequence 7, Application US/09462478A
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND
; TITLE OF INVENTION: SUBLANCIN LANTIBIOTIC PRODUCED BY BACILLUS SUBTILIS 168
; FILE REFERENCE: 8172-8072
; CURRENT APPLICATION NUMBER: US/09/462,478A
; CURRENT FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: US/60/053,035
; PRIOR FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Bacillus subtilis 168
; FEATURE:
; OTHER INFORMATION: Prosublancin 168
US-09-462-478A-7

Query Match 38.7%; Score 211; DB 18; Length 37;
Best Local Similarity 100.0%; Pred. No. 9.3e-15;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 GLGKAQCAALWLCASGGTIGCGGAVACQNYRQFCR 56
|||||
Db 1 GLGKAQCAALWLCASGGTIGCGGAVACQNYRQFCR 37
|||||

RESULT 10
PCT-US98-14547-18
; Sequence 18, Application PC/TUS9814547B
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND
; TITLE OF INVENTION: SUBLANCIN LANTIBIOTIC PRODUCED BY BACILLUS SUBTILIS 168
; FILE REFERENCE: 8172-8068
; CURRENT APPLICATION NUMBER: PCT/US98/14547B
; CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: US/60/053,035
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Bacillus subtilis 168
; FEATURE:
; NAME/KEY: DISULFID
; LOCATION: (14)..(29)
; FEATURE:
; NAME/KEY: DISULFID
; LOCATION: (7)..(36)
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (16)
; OTHER INFORMATION: Xaa is a dehydrogenated Ser (Dha)
; FEATURE:
; NAME/KEY: THIOETH
; LOCATION: (19)..(22)
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (19)
; OTHER INFORMATION: Xaa is a dehydrogenated Thr (Dhb)
PCT-US98-14547-18

Query Match 37.1%; Score 202; DB 1; Length 37;

```
; NAME/KEY: DISULFID
; LOCATION: (7)..(36)
; NAME/KEY: MOD_RES
; LOCATION: (16)
; OTHER INFORMATION: xaa is a dehydrogenated Ser (Dha)
; NAME/KEY: THIOETH
; LOCATION: (19)..(22)
; NAME/KEY: MOD_RES
; LOCATION: (19)
; OTHER INFORMATION: Xaa is a dehydrogenated Thr (Dhb)
US-09-462-478A-18

Query Match          37.1%; Score 202; DB 18; Length 37;
Best Local Similarity 94.6%; Pred. No. 8.4e-14;
Matches 35; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY   20 GLGKAQCAALWLQCASGGTIGCGGGAVACQNYRQFCR 56
      |||||
Db    1 GLGKAQCAALWLQCAXGGXIGCGGGAVACQNYRQFCR 37

RESULT 13
US-08-085-126-56
; Sequence 56, Application US/08085126
; GENERAL INFORMATION:
; APPLICANT: LADNER, Robert C.
; APPLICANT: CANNON, Larick E.
; TITLE OF INVENTION: CHELYDRYZYMES AND NEPAZYMES, AND USES
; TITLE OF INVENTION: THEREFOR
; NUMBER OF SEQUENCES: 236
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/085,126
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: LADNER=10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID`NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-085-126-56

Query Match          26.4%; Score 144; DB 4; Length 396;
Best Local Similarity 53.1%; Pred. No. 1.6e-06;
Matches 26; Conservative 6; Mismatches 17; Indels 0; Gaps 0;

QY   36 GGTTICGGGAVACQNYRQFCRGGGGGGGGGGGGGSKFD DFD 84
      || : | | : : : | | | | | | | | | | | |
Db    104 GGC MGNNFKSAEDCMRTCGAGGGGGGGGGGGGAAAPSF 152

RESULT 14
US-08-438-114-56
```



```

; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: 5202-D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 498-8260
; TELEFAX: 617 876-5851
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-808-324-30

Query Match      24.9%; Score 135.5; DB 4; Length 240;
Best Local Similarity 40.4%; Pred. No. 7.2e-07;
Matches 36; Conservative 5; Mismatches 25; Indels 23; Gaps 4;

QY 2 EKLFEVKLE-----ELENQKSGGLG--KAQCAALWLQ-CASGGTIGCGGAVACQN 50
   |||:::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 53 ESLFREIRAQARALRAAAEPPDPGAGSRKANLGGRRRQRTALAGTRGXGSGS----- 106
   |||:::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

QY 51 YRQFCRGGGGGGGGGGGGGGGGGGGMSK 79
   |||:::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 107 -----GGGGGGGGGGGGGGGGGGGAGR 129
   |||:::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 5
PCT-US94-14030A-30
; Sequence 30, Application PC/TUS9414030A
; GENERAL INFORMATION:
; APPLICANT: GENETICS INSTITUTE, INC.
; APPLICANT: PRESIDENT AND FELLOWS OF HARVARD COLLEGE
; TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/14030A
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/164,103
; FILING DATE: 07-DEC-1993
; APPLICATION NUMBER: US 08/217,780
; FILING DATE: 25-MAR-1994
; APPLICATION NUMBER: US 08/333,576
; FILING DATE: 02-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: 5202D-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 498-8260
; TELEFAX: 617 876-5851
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

```

PCT-US94-14030A-30

Query Match      24.9%; Score 135.5; DB 5; Length 240;
Best Local Similarity 40.4%; Pred. No. 7.2e-07;
Matches 36; Conservative 5; Mismatches 25; Indels 23; Gaps 4;

QY 2 EKLFEVKLE-----ELENQKSGGLG--KAQCAALWLQ-CASGGTIGCGGAVACQN 50
   |||:::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 53 ESLFREIRAQARALRAAAEPPDPGAGSRKANLGGRRRQRTALAGTRGXGSGS----- 106
   |||:::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

QY 51 YRQFCRGGGGGGGGGGGGGGGGGGGMSK 79
   |||:::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 107 -----GGGGGGGGGGGGGGGGGGGAGR 129
   |||:::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 6
US-09-052-995-1
; Sequence 1, Application US/09052995
; Patent No. 6183956
; GENERAL INFORMATION:
; APPLICANT: Sivaraja, Mohanram
; APPLICANT: Strulovici, Berta
; APPLICANT: Flores, Osvaldo A.
; TITLE OF INVENTION: High Throughput In Vitro Screening Assay
; TITLE OF INVENTION: for Transcription Modulators
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/052,995
; FILING DATE: 31-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parent, Annette S.
; REGISTRATION NUMBER: 42,058
; REFERENCE/DOCKET NUMBER: 018781-0006000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-052-995-1

Query Match      24.3%; Score 132.5; DB 4; Length 201;
Best Local Similarity 63.4%; Pred. No. 1.2e-06;
Matches 26; Conservative 0; Mismatches 4; Indels 11; Gaps 1;

QY 36 GGTICGGGAVACQNYRQFCRGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 76
   |||:::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 99 GPGGGGGGGG-----GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 128
   |||:::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 7
US-09-053-003-40
; Sequence 40, Application US/09053003
; Patent No. 6207391
; GENERAL INFORMATION:
; APPLICANT: Wu, Pengguang

```

```

; APPLICANT: McKinney, Judi
; TITLE OF INVENTION: High-Throughput Screening Assays for
; TITLE OF INVENTION: Modulators of STAT4 and STAT6 Activity
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/053,003
; FILING DATE: 31-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Timothy L.
; REGISTRATION NUMBER: 35,367
; REFERENCE/DOCKET NUMBER: 018781-0008000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1..97
; OTHER INFORMATION: /product= "OTHER"
; OTHER INFORMATION: /note= "Gly at positions 1-97 may be
; OTHER INFORMATION: present or absent"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 105..201
; OTHER INFORMATION: /product= "OTHER"
; OTHER INFORMATION: /note= "Gly at positions 105-201 may be
; OTHER INFORMATION: present or absent"
US-09-053-003-40

```

```

Query Match 24.3%; Score 132.5; DB 4; Length 201;
Best Local Similarity 63.4%; Pred. No. 1.2e-06;
Matches 26; Conservative 0; Mismatches 4; Indels 11; Gaps 1;

```

```

QY 36 GGTIGCGGAVACQNYRQFCRGGGGGGGGGGGGGGGGGGGG 76
Db 99 GGPGGGGGGG-----GGGGGGGGGGGGGGGGGGGG 128

```

```

RESULT 8
US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; TITLE OF INVENTION: TO SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isshinden
; CITY: Tsu-city

```

```

; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

```

```

Query Match 23.8%; Score 129.5; DB 3; Length 738;
Best Local Similarity 41.0%; Pred. No. 9.2e-06;
Matches 32; Conservative 3; Mismatches 12; Indels 31; Gaps 3;

```

```

QY 18 GSGLGKAQCAALWLQCASGGTIGCGGAVACQNYRQFCRGGGGGGGGGGGGGGGGGGGG-- 75
Db 674 GSGGG-----GSGGGGGGGGGG-----GGGGGGGGGGGGGGGGGGGG 711

```

```

QY 76 -----GMSKFFDDFDLD 86
Db 712 NNGWNGNNGNKKYDDDDCD 729

```

```

RESULT 9
US-08-835-099A-9
; Sequence 9, Application US/08835099A
; Patent No. 5874277
; GENERAL INFORMATION:
; APPLICANT: SHINTANI, Yasushi
; APPLICANT: NISHI, Kazunori
; APPLICANT: KAWAMOTO, Tomohiro
; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
; TITLE OF INVENTION: AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:

```


ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 47342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440

RESULT 12

```

PCT-US94-07799-6
; Sequence 6, Application PC/TUS9407799
; GENERAL INFORMATION:
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE
; TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-7
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spensley Horn Jubas & Lubitz
; STREET: 1880 Century Park East, Suite 500
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07799
; FILING DATE: 08-JUL-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: TUMARKIN, LISA A., PH.D.
; REGISTRATION NUMBER: P-38,347
; REFERENCE/DOCKET NUMBER: FD-2348
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-07799-6

```

Query Match	22.9%	Score 125;	DB 5;	Length 161;
Best Local Similarity	54.3%	Pred. No. 5.8e-06;		
Matches 25; Conservative	1;	Mismatches 8;	Indels 12;	Gaps 1;
OY	34	AGGTTICGGGAVACQNYRQFCRGGGGGGGGGGGGGMSK	79	
	17	ALACTRGAQGS-----GGGGGGGGGGGGGGGGGAGR	50	
Db				

```

RESULT 13
US-09-032-523-7
; Sequence 7, Application US/09032523
; Patent No. 6232454
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl
; APPLICANT: Baugh, Mariah
; TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032,523
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0479 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 266 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 164403
; US-09-032-523-7

Query Match 22.9%; Score 125
Best Local Similarity 44.6%; Pred. No.
Matches 29; Conservative 3; Mismatch

Qy 17 KGSGLGKAQCAALWLQCASGGTIGCG-----GGG
||| || | | | | | | | | | | | | |
Db 9 KGGGGGG-----GGGGGLGGGLGNVLGG

QY 72 GGGGG 76
|||||
Db 50 GGGGG 54

RESULT 14
US-08-014-943A-2
; Sequence 2, Application US/08014943A
; Patent No. 5545551
; GENERAL INFORMATION:
; APPLICANT: Johnson, Edward M.
; APPLICANT: Bergemann, Andrew D.
; TITLE OF INVENTION: Cloning And Expre
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Ve
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/014,943A
; FILING DATE: 02/FEB/1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 322 amino acids
; TYPE: amino acid

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:26:16 ; Search time 18.9781 Seconds
(without alignments)
506.317 Million cell updates/sec

Title: US-09-894-030-3
Perfect score: 545
Sequence: 1 MEKLFKEVKLELENQKGS.....DDFDLDVVKVSKQDSKITPQ 100

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	304	55.8	56	T12783	sublancin 168 prec
2	165	30.3	336	S18750	chitinase (EC 3.2.
3	151	27.7	385	T20410	hypothetical prote
4	142.5	26.1	919	A39248	androgen receptor
5	141	25.9	221	T04592	glycine-rich cell
6	139.5	25.6	393	T20268	hypothetical prote
7	138	25.3	207	T07381	glycine-rich prote
8	134.5	24.7	1084	T04103	sucrose-phosphate
9	132.5	24.3	433	S20963	homeotic protein H
10	130	23.9	165	1 KNRZG1	glycine-rich cell
11	130	23.9	255	B84777	hypothetical prote
12	130	23.9	291	S31415	glycine-rich prote
13	129	23.7	106	F84797	hypothetical prote
14	129	23.7	268	1 CIHUL	calpain (EC 3.4.22
15	129	23.7	431	1 WJHU2G	homeotic protein H
16	128	23.5	214	1 KNNT2S	glycine-rich prote
17	128	23.5	681	2 AB2155	hypothetical prote
18	127	23.3	271	S34666	glycine-rich prote
19	127	23.3	384	1 A26099	glycine-rich cell
20	127	23.3	396	2 T49109	glycine-rich prote
21	126	23.1	290	2 AD1849	hypothetical prote
22	126	23.1	333	2 A39065	homeotic protein E
23	125.5	23.0	404	2 S54729	RNA-binding protei
24	125.5	23.0	592	2 E82759	endo-1,4-beta-gluc
25	125	22.9	151	2 S43296	bone morphogenetic
26	125	22.9	220	2 A44805	eggshell protein p
27	125	22.9	266	1 CIPGL	calpain (EC 3.4.22
28	124.5	22.8	321	2 A38712	fibrillarin [valid
29	124	22.8	201	2 F84596	glycine-rich prote

30	124	22.8	322	2	A45036	single-stranded-DN
31	123.5	22.7	152	2	G96010	hypothetical expor
32	123	22.6	280	2	A42424	chitinase (EC 3.2.
33	123	22.6	1585	2	T31611	hypothetical prote
34	122.5	22.5	136	2	T29282	hypothetical prote
35	122	22.4	388	2	T29173	hypothetical prote
36	121.5	22.3	171	2	H84709	probable glycine-r
37	121.5	22.3	268	2	S09860	hypothetical prote
38	121.5	22.3	302	2	C84470	hypothetical prote
39	121.5	22.3	316	1	A38743	loricrin - human
40	121.5	22.3	481	2	A35628	loricrin - mouse
41	121.5	22.3	549	2	B86264	hypothetical prote
42	121	22.2	56	1	NIBSSA	subtilin precursor
43	120.5	22.1	199	2	S16063	acp-22 protein - y
44	120.5	22.1	199	2	S32224	acp-22 protein - y
45	120.5	22.1	266	1	CIRBL	calpain (EC 3.4.22

ALIGNMENTS

RESULT 1
T12783
sublancin 168 precursor - Bacillus subtilis phage SPBc2
C;Species: Bacillus subtilis phage SPBc2
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 24-Sep-1999
C;Accession: T12783; H69719
R;Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Mauel, C.; Karamata, D.
submitted to the EMBL Data Library, August 1997
A;Description: The complete nucleotide sequence of the Bacillus subtilis Spbetac2 pro
A;Reference number: Z17583
A;Accession: T12783
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-56 <LAZ>
A;Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025497; PIDN:AAC12992.1
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, F.; Sekiguchi, J.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033
A;Accession: H69719
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-56 <KUN>
A;Cross-references: GB:299115; GB:AL009126; NID:g2634478; PIDN:CAB14066.1; PID:el1835
A;Experimental source: strain 168
C;Genetics: <LA1>
C;Gene: yolG
C;Genetics: <KUL>
A;Gene: sunA
C;Superfamily: unassigned lanthionine-containing peptides
C;Keywords: antibiotic; lanthionine
F;1-19/Domain: propeptide #status predicted <PRO>
F;20-56/Product: sublancin 168 #status predicted <MAT>
F;26-55/Disulfide bonds: #status experimental
F;33-48/Disulfide bonds: #status predicted
F;35/Modified site: dehydroalanine (Ser) #status experimental
F;38-41/Cross-link: (2S,3S,6R)-3-methyl-lanthionine (Thr-Cys) #status predicted

Query Match 55.8%; Score 304; DB 2; Length 56;
Best Local Similarity 100.0%; Pred. No. 4.4e-22;


```

RESULT 6
T20268
hypothetical protein C56A3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20268
R:Sims, M.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z19244
A:Accession: T20268
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-393 <WIL>
A:Cross-references: EMBL:Z77655; PIDN:CAB01137.1; GSPDB:GN00023; CESP:C56A3
A:Experimental source: clone C56A3
C:Genetics:
A:Gene: CESP:C56A3.1
A:Map position: 5
A:Introns: 51/3; 91/1; 121/1; 331/3

Query Match      25.6%; Score 139.5; DB 2; Length 393;
Best Local Similarity 50.8%; Pred. No. 7e-06;
Matches 30; Conservative 0; Mismatches 16; Indels 13; Gaps

QY 18 GSGLGKQAQCAALWLQCASGGTIGCGGGAVACQNYRQFCRGGGGGGGGGGGGGGG 76
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87 GGGCGGGGCGGGGGCGGGG--GCGGGG-----GGGCGGGGGGGCGGGGGGGG 132

RESULT 7
T07381
glycine-rich protein Tfm5 - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
C:Accession: T07381
R:Santino, C.G.; Stanford, G.L.; Conner, T.W.
Plant Mol. Biol. 33, 405-416, 1997
A:Title: Developmental and transgenic analysis of two tomato fruit enhanced
A:Reference number: Z16000; MUID:97201476
A:Accession: T07381
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-207 <SAN>
A:Cross-references: EMBL:X95262; NID:g1166449; PIDN:CAA64559.1; PID:g1166445
A:Experimental source: cultivar UC82b; fruit
C:Genetics:
A:Gene: Tfm5
C:Superfamily: hydroxyproline-rich glycoprotein

Query Match      25.3%; Score 138; DB 2; Length 207;
Best Local Similarity 50.8%; Pred. No. 5.7e-06;
Matches 30; Conservative 1; Mismatches 6; Indels 22; Gaps

QY 18 GSGLGKQAQCAALWLQCASGGTIGCGGGAVACQNYRQFCRGGGGGGGGGGGGGGG 76
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 85 GSGSG-----GGSGGTGGGGG-----GGGGGGGGGGGGGGGGGGG 121

RESULT 8
T04103
sucrose-phosphate synthase (EC 2.4.1.14) 1 - rice
C:Species: Oryza sativa (rice)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
C:Accession: T04103
R:Sakamoto, M.; Satozawa, T.; Kishimoto, N.; Higo, K.; Shimada, H.; Fujimura
Plant Sci. 112, 207-217, 1995
A:Title: Structure and RFLP mapping of a rice sucrose phosphate synthase (S
A:Reference number: Z15212
A:Accession: T04103
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA

```


A; Gene: GDB:CAPN4

Thu Oct 24 14:40:01 2002

us-09-894-030-3.rpr

Page 6

Search completed: October 24, 2002, 09:29:44
Job time : 20.9781 secs

; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
 ; FILE REFERENCE: 70030-NP
 ; CURRENT APPLICATION NUMBER: US/10/259,165
 ; CURRENT FILING DATE: 2002-09-26
 ; PRIOR APPLICATION NUMBER: US 60/370,620
 ; PRIOR FILING DATE: 2002-04-04
 ; PRIOR APPLICATION NUMBER: US 60/368,327
 ; PRIOR FILING DATE: 2002-03-27
 ; PRIOR APPLICATION NUMBER: US 60/325,277
 ; PRIOR FILING DATE: 2001-09-26
 ; NUMBER OF SEQ ID NOS: 782
 ; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
 ; SEQ ID NO 212
 ; LENGTH: 280
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 US-10-259-165-212

Query Match 26.0%; Score 141.5; DB 6; Length 280;
 Best Local Similarity 46.2%; Pred. No. 3.4e-05;
 Matches 30; Conservative 3; Mismatches 15; Indels 17; Gaps 3;

QY 20 GLGKAQCAALWLCASGGTIGCGGAVACQNYRQFCR-----GGGGGGGGGGGGGGG 73
 Db 25 GCQNMCCSKWGYCGTKDY-CGDG-----CRSGPCYGGGGGGGGGGGGGGG 73

QY 74 GGGMS 78
 Db 74 GSGVS 78

RESULT 11
 US-09-791-537-78475
 ; Sequence 78475, Application US/09791537
 ; GENERAL INFORMATION:
 ; APPLICANT: Bionomix, Inc.
 ; APPLICANT: Debe, Derek
 ; APPLICANT: Danzer, Joseph
 ; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE
 ; FILE REFERENCE: 261/210
 ; CURRENT APPLICATION NUMBER: US/09/791,537
 ; CURRENT FILING DATE: 2001-02-22
 ; NUMBER OF SEQ ID NOS: 153055
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 78475
 ; LENGTH: 318
 ; TYPE: PRT
 ; ORGANISM: Allium sativum
 US-09-791-537-78475

Query Match 25.9%; Score 141; DB 5; Length 318;
 Best Local Similarity 46.2%; Pred. No. 4.3e-05;
 Matches 37; Conservative 5; Mismatches 22; Indels 16; Gaps 5;

QY 4 LPKEVKLEENQKSGGLGKAQCAALWLCASGGTIG-----CGGGAVACQNYRQFCRGG 58
 Db 11 LFK-----NSYAQCQGSQAGGALCSNR-LCCSRFGYCGSTDPYCGTG---CQSQ---CGGG 59
 QY 59 GGGGGGGGGGGGGGGGMS 78
 Db 60 GGGGGGGGGGGGGGSGVA 79

RESULT 12
 US-09-791-537-22910
 ; Sequence 22910, Application US/09791537
 ; GENERAL INFORMATION:
 ; APPLICANT: Bionomix, Inc.
 ; APPLICANT: Debe, Derek
 ; APPLICANT: Danzer, Joseph
 ; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE
 ; FILE REFERENCE: 261/210
 ; CURRENT APPLICATION NUMBER: US/09/791,537
 ; CURRENT FILING DATE: 2001-02-22
 ; NUMBER OF SEQ ID NOS: 153055
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 78475
 ; LENGTH: 318
 ; TYPE: PRT
 ; ORGANISM: Allium sativum
 US-09-791-537-78475

; FILE REFERENCE: 261/210
 ; CURRENT APPLICATION NUMBER: US/09/791,537
 ; CURRENT FILING DATE: 2001-02-22
 ; NUMBER OF SEQ ID NOS: 153055
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 22910
 ; LENGTH: 335
 ; TYPE: PRT
 ; ORGANISM: Elaeagnus umbellata
 US-09-791-537-22910

Query Match 25.5%; Score 139; DB 5; Length 335;
 Best Local Similarity 51.6%; Pred. No. 6.8e-05;
 Matches 33; Conservative 3; Mismatches 16; Indels 12; Gaps 4;

QY 16 QKSGGLGKAQCAALWLCASGGTIG-----CGGGAVACQNYRQFCRGGGGGGGGGGG 70
 Db 24 QCGQQLGGALCSG-GLCCSQWGYCGNTDPYCGDG---CQSQ---CDGGSGGGGGGGGGG 76

QY 71 GGGG 74
 Db 77 GGGG 80

RESULT 13
 US-09-945-182-30
 ; Sequence 30, Application US/09945182
 ; GENERAL INFORMATION:
 ; APPLICANT: Celeste, Anthony J.
 ; APPLICANT: Wozney, John
 ; APPLICANT: Rosen, Vicki A.
 ; APPLICANT: Wolfman, Neil
 ; APPLICANT: Thomsen, Gerald H.
 ; APPLICANT: Melton, Douglas A.
 ; TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
 ; NUMBER OF SEQUENCES: 35
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: GENETICS INSTITUTE, INC.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/945,182
 ; FILING DATE: 31-Aug-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/808,324
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lazar, Steven R.
 ; REGISTRATION NUMBER: 32,618
 ; REFERENCE/DOCKET NUMBER: 5202-D
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617 498-8260
 ; TELEFAX: 617 876-5851
 ; INFORMATION FOR SEQ ID NO: 30:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 240 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 30:
 US-09-945-182-30

Query Match 24.9%; Score 135.5; DB 5; Length 240;
 Best Local Similarity 40.4%; Pred. No. 0.0001;

Thu Oct 24 14:40:00 2002

	Matches	36;	Conservative	5;	Mismatches	25;	Indels	23;	Gaps	4;
QY	2	EKL	FKEVKLE-----	ELENQK	GSLG--	KAQCAAL	WLQ-CASGGT	ICGGG	AVACON	50
			: : :							
Dd	53	ESL	FEIRIAQARALRAAAE	PPDP	PGGAGSRKANL	GRRRQRTALAG	TRGXXGS-----			106
QY	51	YRQ	FCRGGGGGGGGGGGGGGGGGG	GMSK	79					
				:						
Dd	107	-----	GGGGGGGGGGGGGGGGGGG	GAGR	129					

```

RESULT 14
US-09-791-537-88976
; Sequence 88976, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 88976
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-791-537-88976

```

	Query Match	24.8%;	Score 135;	DB 5;	Length 331;
	Best Local Similarity	38.9%;	Pred. No.	0.00015;	
	Matches	28;	Conservative	11;	Mismatches 19; Indels 14; Gaps 2;
QY	34	ASGGTTCGCGGAVACQNYRQFCRCGGGGGGGGGGGGGGGGGGGGGMSKFFDDFDLDVVVKVSKQ	93		
		::	: :	:	:::
Dd	57	AAGGGGAGGGGGG-----GGGGGGGAGGGAGGGGRSPVRELDMGAAERSRE	107		
QY	94	D-----SKITPQ	100		
		:::: :			
Dd	108	PGSPRLTEVSPE	119		

RESULT 15
US-10-182-995-20071
; Sequence 20071, Application US/10182995
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HUMAN HEART
; FILE REFERENCE: PB 0004 WO 1
; CURRENT APPLICATION NUMBER: US/10/182,995
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 29119

```

; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 20071
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000161.1
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
US-10-182-995-20071

```

	Query Match	24.7%	Score 134.5;	DB 6;	Length 101;
	Best Local Similarity	44.6%	Pred. No. 5.9e-05;		
	Matches 33;	Conservative	3;	Mismatches 15;	Indels 23; Gaps 3;
QY	6	KEVKLEELNQSGGLGKAQCAALWLQCASGGTIGCGGGAACQNYRQFCRGGGGGGGG	65		
		: :			
Db	11	KRKKEEEEEGREEG-GR-----GGGGGGGGGG	47		
QY	66	GGGGGGGGGGMSK	79		
Db	48	GGGGGGGGGRRRK	61		

Search completed: October 24, 2002, 09:35:11
Job time : 58.2044 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:24:01 ; Search time 11.6788 Seconds
(without alignments)
331.536 Million cell updates/sec

Title: US-09-894-030-3
Perfect score: 545
Sequence: 1 MEKLFKEVKLEENKGGSG.....DDFDLDVVKVSKQDSKITPQ 100

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	165	30.3	303	1 CHIB_POPTR	P29031 populus tri
2	142.5	26.1	919	1 ANDR_HUMAN	P10275 homo sapien
3	135	24.8	331	1 SHX2_MOUSE	P70390 mus musculu
4	132.5	24.3	433	1 HXB3_MOUSE	P09026 mus musculu
5	132	24.2	331	1 SHX2_HUMAN	O60902 homo sapien
6	130	23.9	165	1 GRP1_ORYSA	P25074 oryza sativ
7	129	23.7	268	1 CANS_HUMAN	P04632 homo sapien
8	129	23.7	431	1 HXB3_HUMAN	P14651 homo sapien
9	128	23.5	214	1 GRP2_NICSY	P27484 nicotiana s
10	127	23.3	384	1 GRP1_PETHY	P09789 petunia hyb
11	126.5	23.2	911	1 ANDR_PANTR	O97775 pan troglod
12	126	23.1	416	1 R23B_MOUSE	P54728 mus musculu
13	126	23.1	476	1 EVX2_HUMAN	Q03828 homo sapien
14	125.5	23.0	404	1 CAZ_DROME	Q27294 drosophila
15	125	22.9	151	1 GDF7_MOUSE	P43029 mus musculu
16	125	22.9	266	1 CANS_PIG	P04574 sus scrofa
17	124.5	22.8	321	1 FBRL_HUMAN	P22087 homo sapien
18	124	22.8	201	1 GR2B_ARATH	Q38896 arabidopsis
19	124	22.8	322	1 PUR_HUMAN	Q00577 homo sapien
20	124	22.8	1627	1 TP2B_CHICK	O42131 gallus gall
21	123.5	22.7	321	1 PUR_MOUSE	P42669 mus musculu
22	123	22.6	280	1 CHIA_MAIZE	P29022 zea mays (m
23	122	22.4	520	1 WASP_MOUSE	P70315 mus musculu
24	121.5	22.3	268	1 EP34_HCMVA	P16768 human cytom
25	121.5	22.3	316	1 LORI_HUMAN	P23490 homo sapien
26	121.5	22.3	481	1 LORI_MOUSE	P18165 mus musculu
27	121.5	22.3	684	1 EP84_HCMVA	P17151 human cytom
28	121	22.2	56	1 LANS_BACSU	P10946 bacillus su
29	120.5	22.1	199	1 AC22_TENMO	P26968 tenebrio mo
30	120.5	22.1	266	1 CANS_RABIT	P06813 oryctolagus
31	120.5	22.1	332	1 SIX3_HUMAN	O95343 homo sapien
32	120.5	22.1	440	1 FXGA_CHICK	Q98937 gallus gall
33	120.5	22.1	458	1 K1CM_HUMAN	P13646 homo sapien

34	120	22.0	443	1 OC3N_HUMAN	P20265 homo sapien
35	120	22.0	445	1 OC3N_MOUSE	P31360 mus musculu
36	120	22.0	533	1 ZIC2_HUMAN	O95409 homo sapien
37	119.5	21.9	252	1 GRP1_PHAVU	P10495 phaseolus v
38	119.5	21.9	694	1 FZD8_HUMAN	Q9h461 homo sapien
39	119.5	21.9	1380	1 DDX9_MOUSE	O70133 mus musculu
40	119	21.8	206	1 TWST_MOUSE	P26687 mus musculu
41	119	21.8	495	1 BRN1_MOUSE	P31361 mus musculu
42	119	21.8	497	1 BRN1_RAT	Q63262 rattus norv
43	119	21.8	500	1 BRN1_HUMAN	P20264 homo sapien
44	118	21.7	1093	1 PER_DROWI	Q03297 drosophila
45	117.5	21.6	263	1 CANS_BOVIN	P13135 bos taurus

ALIGNMENTS

RESULT 1
CHIB_POPTR
ID CHIB_POPTR STANDARD; PRT; 303 AA.
AC P29031;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Acidic endochitinase WIN6.2B precursor (EC 3.2.1.14).
OS Populus trichocarpa (Western balsam poplar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Salicaceae; Populus.
OX NCBI_TaxID=3694;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92003678; PubMed=1912489;
RA Davis J.M., Clarke H.R.G., Bradshaw H.D. Jr., Gordon M.P.;
RT "Populus chitinase genes: structure, organization, and similarity of translated sequences to herbaceous plant chitinases.";
RL Plant Mol. Biol. 17:631-639(1991).
CC -!- FUNCTION: THIS PROTEIN FUNCTIONS AS A DEFENSE AGAINST CHITIN CONTAINING FUNGAL PATHOGENS.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-acetyl-D-glucosamine polymers of chitin.
CC -!- INDUCTION: By wounding.
CC -!- SIMILARITY: BELONGS TO CHITINASE CLASS IA WHICH IS SIMILAR TO CLASS IB IN THE CATALYTIC PART BUT WHICH INCLUDE A N-TERMINAL CHITIN-BINDING LECTIN DOMAIN (BELONGS TO FAMILY 19 OF GLYCOSYL HYDROLASES).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL; X59995; CAA42612.1; -.
DR HSSP; P02877; IHEV.
DR InterPro; IPR001002; Chitin_bind.
DR InterPro; IPR000726; Glyco_hydro_19.
DR Pfam; PF00187; chitin_binding; 1.
DR Pfam; PF00182; Glyco_hydro_19; 2.
DR PRINTS; PR00451; CHITINBINDNG.
DR ProDom; PD000574; Glyco_hydro_19; 1.
DR ProDom; PD000609; Chitin_bind; 1.
DR SMART; SM00270; ChtBD1; 1.
DR PROSITE; PS00026; CHITIN_BINDING; 1.
DR PROSITE; PS00773; CHITINASE_19_1; 1.
DR PROSITE; PS00774; CHITINASE_19_2; 1.
KW Hydrolase; Glycosidase; Chitin degradation; Chitin-binding; Signal;
KW Multigene family.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 303 ACIDIC ENDOCHITINASE WIN6.2B.
FT DOMAIN 22 61 CHITIN-BINDING (BY SIMILARITY).

RA Muroso K., Zhou Z.;
RT "Molecular genetics of human androgen insensitivity.";
RL Eur. J. Pediatr. 152 Suppl. 2:S62-S69(1993).
RN [16]
RP REVIEW ON VARIANTS.
RX MEDLINE=94059770; PubMed=8240973;
RA Sultan C., Lumbroso S., Poujol N., Belon C., Boudon C.,
RA Lobaccaro J.-M.;
RT "Mutations of androgen receptor gene in androgen insensitivity
RT syndromes.";
RL J. Steroid Biochem. Mol. Biol. 46:519-530(1993).
RN [17]
RP REVIEW ON VARIANTS.
RX MEDLINE=95023089; PubMed=7937057;
RA Patterson M.N., Hughes I.A., Gottlieb B., Pinsky L.;
RT "The androgen receptor gene mutations database.";
RL Nucleic Acids Res. 22:3560-3562(1994).
RN [18]
RP REVIEW ON VARIANTS.
RX MEDLINE=95352489; PubMed=7626493;
RA Brinkmann A.O., Jenster G., Ris-Stalpers C., van der Korput J.A.G.M.,
RA Bruggenwirth H.T., Boehmer A.L.M., Trapman J.;
RT "Androgen receptor mutations.";
RL J. Steroid Biochem. Mol. Biol. 53:443-448(1995).
RN [19]
RP REVIEW ON VARIANTS.
RX MEDLINE=97169385; PubMed=9016528;
RA Gottlieb B., Trifiro M.A., Lumbroso R., Vasilidou D.M., Pinsky L.;
RT "The androgen receptor gene mutations database.";
RL Nucleic Acids Res. 25:158-162(1997).
RN [20]
RP VARIANT LNCAP ALA-877.
RX MEDLINE=91083633; PubMed=2260966;
RA Veldscholte J., Ris-Stalpers C., Kuiper G.G.J.M., Jenster G.,
RA Berrevoets C.A., Claassen E., van Rooij H.C.J., Trapman J.,
RA Brinkmann A.O., Mulder E.;
RT "A mutation in the ligand binding domain of the androgen receptor of
RT human LNCap cells affects steroid binding characteristics and
RT response to anti-androgens.";
RL Biochem. Biophys. Res. Commun. 173:534-540(1990).
RN [21]
RP VARIANTS CAIS CYS-774.
RX MEDLINE=91186983; PubMed=2082179;
RA Brown T.R., Lubahn D.B., Wilson E.M., French F.S., Migeon C.J.,
RA Corfen J.L.;
RT "Functional characterization of naturally occurring mutant androgen
RT receptors from subjects with complete androgen insensitivity.";
RL Mol. Endocrinol. 4:1759-1772(1990).
RN [22]
RP VARIANT CYS-774.
RX MEDLINE=91310758; PubMed=1856263;
RA Marcelli M., Tilley W.D., Zoppi S., Griffin J.E., Wilson J.D.,
RA McPhaul M.J.;
RT "Androgen resistance associated with a mutation of the androgen
RT receptor at amino acid 772 (Arg-->Cys) results from a combination of
RT decreased messenger ribonucleic acid levels and impairment of
RT receptor function.";
RL J. Clin. Endocrinol. Metab. 73:318-325(1991).
RN [23]
RP VARIANT CAIS PRO-617.
RX MEDLINE=91154385; PubMed=1999491;
RA Marcelli M., Zoppi S., Grino P.B., Griffin J.E., Wilson J.D.,
RA McPhaul M.J.;
RT "A mutation in the DNA-binding domain of the androgen receptor gene
RT causes complete testicular feminization in a patient with
RT receptor-positive androgen resistance.";
RL J. Clin. Invest. 87:1123-1126(1991).
RN [24]
RP VARIANT PAIS CYS-763.
RX MEDLINE=91185626; PubMed=2010552;
RA McPhaul M.J., Marcelli M., Tilley W.D., Griffin J.E.,
RA Isidro-Gutierrez R.F., Wilson J.D.;
RT "Molecular basis of androgen resistance in a family with a qualitative

RT abnormality of the androgen receptor and responsive to high-dose
RT androgen therapy.";
RL J. Clin. Invest. 87:1413-1421(1991).
RN [25]
Query Match 26.1%; Score 142.5; DB 1; Length 919;
Best Local Similarity 42.5%; Pred. No. 7.7e-06;
Matches 37; Conservative 3; Mismatches 22; Indels 25; Gaps 4;
QY 12 ELENQKSGGLGKAQCAALWLQCASG-----GTIGCGG--AVACQNYRQF----- 54
Db 386 KLENPLDYGSAAAAA---QCRYGDLASLHGAGAAGPGSGSPSAAASSSWHTLFTAEEG 442
QY 55 -----CRGGGGGGGGGGGGGGGGGG 76
Db 443 QLYGPCGGGGGGGGGGGGGGGGGG 469
RESULT 3
SHX2_MOUSE
ID SHX2_MOUSE STANDARD; PRT; 331 AA.
AC P70390; P70369;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Short stature homeobox protein 2 (Homeobox protein Ogl2X) (OG-12)
DE (Paired family homeodomain protein Prx3).
GN SHOX2 OR Ogl2X OR PRX3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Embryo;
RX MEDLINE=98058757; PubMed=9371788;
RA van Schaick H.S.A., Smidt M.P., Rovescalli A.C., Luijten M.,
RA van der Kleij A.A.M., Asoh S., Kozak C.A., Nirenberg M.W.,
RA Burbach J.P.H.;
RT "Homeobox gene Prx3 expression in rodent brain and extraneural
RT tissues.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:12993-12998(1997).
RN [2]
RP SEQUENCE OF 116-331 AND 235-331 FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=BALB/c; TISSUE=Liver, and Embryo;
RX MEDLINE=97008065; PubMed=8855241;
RA Rovescalli A.C., Asoh S., Nirenberg M.W.;
RT "Cloning and characterization of four murine homeobox genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:10691-10696(1996).
RN [3]
RP DEVELOPMENTAL EXPRESSION.
RX MEDLINE=98133920; PubMed=9466998;
RA Semina E.V., Reiter R.S., Murray J.C.;
RT "A new human homeobox gene Ogl2X is a member of the most conserved
RT homeobox gene family and is expressed during heart development in
RT mouse.";
RL Hum. Mol. Genet. 7:415-422(1998).
CC -!- FUNCTION: May be a growth regulator and have a role in specifying
CC neural systems involved in processing somatosensory information,
CC as well as in face and body structure formation. May also have a
CC role in heart development.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1/OG12A/PRX3A (shown here) and
CC 2/OG12B/PRX3B; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Highly expressed in striated muscle followed
CC by liver, kidney, testis, brain, heart, lung and spleen.
CC -!- DEVELOPMENTAL STAGE: Expressed from E9 TO E16 day in the heart,
CC otic region, maxillary and mandibular components of the first
CC branchial arch, nasal processes, eyelid, midbrain, medulla
CC oblongata, limbs, dorsal root ganglia and genital tubercle. Also
CC expressed in non-neuronal structures around the oral cavity and in
CC hip and shoulder regions and in mesenchyme surrounding the
CC vertebrae.
CC

Db 142 LKNSPQTAEGCGGG-----GGGGGGGGGGSSGGGGGGGGGDK 183

RESULT 5

SHX2_HUMAN STANDARD; PRT; 331 AA.

AC O60902; O60903; O60465; O60467;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Short stature homeobox protein 2 (Paired-related homeobox protein

DE SHOT) (Homeobox protein Ogl2X).

GN SHOX2 OR SHOT OR OGL2X.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).

RC TISSUE=Fibroblast;

RX MEDLINE=98151525; PubMed=9482898;

RA Blaschke R.J., Monaghan A.P., Schiller S., Schechinger B., Rao E.,

RA Padilla-Nash H., Ried T., Rappold G.A.;

RT "SHOT, a SHOX-related homeobox gene, is implicated in craniofacial,

RT brain, heart, and limb development.";

RL Proc. Natl. Acad. Sci. U.S.A. 95:2406-2411(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Muscle;

RA Strausberg R.;

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 116-331 FROM N.A. (ISOFORM 2).

RC TISSUE=Craniofacial;

RX MEDLINE=98133920; PubMed=9466998;

RA Semina E.V., Reiter R.S., Murray J.C.;

RT "A new human homeobox gene OGL2X is a member of the most conserved

RT homeobox gene family and is expressed during heart development in

RT mouse.";

RL Hum. Mol. Genet. 7:415-422(1998).

CC -!- FUNCTION: May be a growth regulator and have a role in specifying

CC neural systems involved in processing somatosensory information,

CC as well as in face and body structure formation.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1/SHOX2A/SHOTA (shown here) and

CC 2/SHOX2B/SHOTB/OG12XB; are produced by alternative splicing.

CC -!- TISSUE SPECIFICITY: Expressed in heart, skeletal muscle, liver,

CC lung, bone marrow fibroblast, pancreas and placenta.

CC -!- DEVELOPMENTAL STAGE: Expressed during craniofacial development as

CC well as in heart.

CC -!- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.

CC "BICOID" SUBFAMILY.

CC -!- SIMILARITY: CONTAINS 1 OAR DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; AJ002367; CAA05341.1; ALT_INIT.

DR EMBL; AJ002368; CAA05342.1; ALT_INIT.

DR EMBL; BC008829; AAH08829.1; -.

DR EMBL; AF022654; AAC39662.1; ALT_INIT.

DR EMBL; AF023203; AAC39663.1; -.

DR HSSP; P06601; 1FJL.

DR MIM; 602504; -.

DR InterPro; IPR000047; HTH_repressr.

DR InterPro; IPR001356; Homeobox.

DR InterPro; IPR003654; OAR_domain.

DR Pfam; PF00046; homeobox; 1.

DR PRINTS; PR00024; HOMEBOX.

DR PRINTS; PR00031; HTHREPRESSR.

DR SMART; SM00389; HOX; 1.

DR PROSITE; PS00027; HOMEBOX_1; 1.

DR PROSITE; PS00071; HOMEBOX_2; 2.

DR PROSITE; PS00803; OAR; 1.

KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;

KW Alternative splicing.

FT DNA_BIND 140 199 HOMEBOX.

FT DOMAIN 313 326 OAR.

FT DOMAIN 60 86 POLY-GLY.

FT VARSPPLIC 235 246 MISSING (IN ISOFORM SHOX2B).

FT CONFLICT 115 115 E -> EGRKPTKAEVQATLLLPGEAFRL (IN REF.

FT 2).

FT CONFLICT 125 125 E -> D (IN REF. 1; CAA05341).

FT CONFLICT 244 244 P -> S (IN REF. 2).

FT CONFLICT 312 312 D -> N (IN REF. 2 AND 3).

FT CONFLICT 325 325 H -> L (IN REF. 3; AAC39663).

SQ SEQUENCE 331 AA; 34964 MW; 55431B073B3B2250 CRC64;

Query Match 24.2%; Score 132; DB 1; Length 331;

Best Local Similarity 37.3%; Pred. No. 3e-05;

Matches 28; Conservative 12; Mismatches 19; Indels 16; Gaps 2;

QY 31 LQCASGGTIGCGGAVACQNYRQFCRGGGGGGGGGGGGGGGGGGGGGGGMSKFDLDLVVKV 90

Db 56 VRAAGGGGGGGGG-----GGGGGGGGGGGAGGGGGRSPVRELDMGAAER 104

QY 91 SKQD-----SKITPQ 100

Db 105 SREPGSPRLTEVSPE 119

RESULT 6

GRPL_ORYSA STANDARD; PRT; 165 AA.

AC P25074;

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DE Glycine-rich cell wall structural protein 1 precursor.

GN GRP-1.

OS -Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzeae; Oryza.

OX NCBI_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. INDICA-IR36;

RX MEDLINE=91370862; PubMed=1716496;

RA Lei M., Wu R.;

RT "A novel glycine-rich cell wall protein gene in rice.";

RL Plant Mol. Biol. 16:187-198(1991).

CC -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).

CC -!- SUBCELLULAR LOCATION: Cell wall (Potential).

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; X53596; CAA37665.1; -.

DR PIR; S13385; KNR2GL.

DR HSSP; P30129; 4DPV.

KW Cell wall; Structural protein; Repeat; Signal.

FT SIGNAL 1 23 POTENTIAL.

FT CHAIN 24 165 GLYCINE-RICH CELL WALL STRUCTURAL

FT DOMAIN 31 159 PROTEIN 1.

FT GLY-RICH.

```
FT REPEAT 56 62 R2 (TYR-RICH).
FT REPEAT 93 99 R2 (TYR-RICH).
FT REPEAT 132 138 R2 (TYR-RICH).
SQ SEQUENCE 165 AA; 13536 MW; E36CE31C3650AC9A CRC64;

Query Match 23.9%; Score 130; DB 1; Length 165;
Best Local Similarity 48.4%; Pred. No. 2.6e-05;
Matches 31; Conservative 3; Mismatches 22; Indels 8; Gaps 2;

QY 18 GSGLGKQAQCAALWLQCGGTIGCGGAVACQNYRQ-----FCRGGGGGGGGGGGGGG 72
    |||||
    |||||
Db 99 GQNGGAQGGQ---SGGGGGGGGGGGGGGSGSGSGYGYGKGGGGGGGGGGGGGG 155

QY 73 GGGG 76
    |||
Db 156 GGGG 159

RESULT 7
CANS_HUMAN STANDARD; PRT; 268 AA.
AC P04632;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Calcium-dependent protease, small subunit (Calpain regulatory subunit)
DE (Calcium-activated neutral proteinase) (CANP).
GN CAPN4 OR CAPNS1 OR CAPNS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86286563; PubMed=3016651;
RA Ohno S., Emori Y., Suzuki K.;
RT "Nucleotide sequence of a cDNA coding for the small subunit of human
RT calcium-dependent protease."
RL Nucleic Acids Res. 14:5559-5559(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87066759; PubMed=3024120;
RA Miyake S., Emori Y., Suzuki K.;
RT "Gene organization of the small subunit of human calcium-activated
RT neutral protease."
RL Nucleic Acids Res. 14:8805-8817(1986).
RN [3]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Adamson A.W., Burkhardt-Schultz K.,
RA Garcia E., Kyle A., Ramirez M., Stillwagen S., Barnes J.,
RA Danganan L., Bruce R., Quan G., Montgomery M., Ow D.,
RA Kobayashi A., Olsen A.O., Carrano A.V.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin, and Uterus;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX MEDLINE=20105516; PubMed=10639123;
RA Strobl S., Fernandez-Catalan C., Braun M., Huber R., Masumoto H.,
RA Nakagawa K., Irie A., Sorimachi H., Bourenkow G., Bartunik H.,
RA Suzuki K., Bode W.;
RT "The crystal structure of calcium-free human m-calpain suggests an
RT electrostatic switch mechanism for activation by calcium."
Proc. Natl. Acad. Sci. U.S.A. 97:588-592(2000).
CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC catalyze limited proteolysis of substrates involved in
CC cytoskeletal remodelling and signal transduction.
CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC (regulatory) subunit.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
```

```
membrane upon Ca++ binding (By similarity).
-!- SIMILARITY: Contains 4 EF-hand calcium-binding domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC EMBL; X04106; CAA27726.1; -.
CC EMBL; M31511; AAA35646.1; -.
CC EMBL; M31502; AAA35646.1; JOINED.
CC EMBL; M31503; AAA35646.1; JOINED.
CC EMBL; M31504; AAA35646.1; JOINED.
CC EMBL; M31505; AAA35646.1; JOINED.
CC EMBL; M31506; AAA35646.1; JOINED.
CC EMBL; M31507; AAA35646.1; JOINED.
CC EMBL; M31508; AAA35646.1; JOINED.
CC EMBL; M31509; AAA35646.1; JOINED.
CC EMBL; M31510; AAA35646.1; JOINED.
CC EMBL; AD001527; AAB51183.1; -.
CC EMBL; AC002984; AAB81546.1; -.
CC EMBL; BC000592; AAH00592.1; -.
CC EMBL; BC007779; AAH07779.1; -.
CC PIR; A23650; CIHUL.
CC PDB; IKFU; 07-DEC-01.
CC MIM; 114170; -.
CC InterPro; IPR002048; EF-hand.
CC Pfam; PF00036; efhand; 3.
CC PROSITE; PS00018; EF_HAND; 2.
KW Calcium-binding; Repeat; 3D-structure.
FT DOMAIN 1 66 GLY-RICH (HYDROPHOBIC).
FT CA_BIND 96 268 CALCIUM-BINDING.
FT CA_BIND 152 163 EF-HAND 1.
FT CA_BIND 182 193 EF-HAND 2.
FT DOMAIN 217 228 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
FT DOMAIN 247 260 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT DOMAIN 10 26 POLY-GLY.
FT DOMAIN 35 56 POLY-GLY.
FT DOMAIN 78 83 POLY-PRO.
SQ SEQUENCE 268 AA; 28316 MW; 17B87A8E47A90632 CRC64;

Query Match 23.7%; Score 129; DB 1; Length 268;
Best Local Similarity 50.0%; Pred. No. 4.7e-05;
Matches 30; Conservative 0; Mismatches 18; Indels 12; Gaps 1;

QY 17 KSGGLGKQAQCAALWLQCGGTIGCGGAVACQNYRQFCRGGGGGGGGGGGGGGGG 76
    |||||
    |||||
Db 9 KGGGGG-----GGGGGGLGGLGNVLGLISAGGGGGGGGGGGGGGGGG 56

RESULT 8
HXB3_HUMAN STANDARD; PRT; 431 AA.
AC P14651; P17484; O95615;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Homeobox protein Hox-B3 (Hox-2G) (Hox-2.7).
GN HOXB3 OR HOX2G.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90098876; PubMed=2574852;
RA Acampora D., D'Esposito M., Faiella A., Pannese M., Migliaccio E.,
RA Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;
RT "The human HOX gene family."
Nucleic Acids Res. 17:10385-10402(1989).
```


Search completed: October 24, 2002, 09:28:20
Job time : 14.6788 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:24:21 ; Search time 32.1168 Seconds
(without alignments)
538.643 Million cell updates/sec

```

Title: US-09-894-030-3
Perfect score: 545
Sequence: 1 MEKLFKEVKLEENOKGSG.....DDFDLDVVKVSKQDSKITPQ 100

```

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 s
```

```
Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archheap:*
```

SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	304	55.8	56	9	O64033	O64033 bacterioph
2	304	55.8	56	16	O34781	O34781 bacillus su
3	151	27.7	385	5	Q93424	Q93424 caenorhabdi
4	147	27.0	333	10	Q42992	Q42992 oryza sativ
5	142.5	26.1	531	4	Q9BZG5	Q9BZG5 homo sapien
6	142.5	26.1	539	4	Q9NUA2	Q9NUA2 homo sapien
7	142.5	26.1	542	4	Q9BZG6	Q9BZG6 homo sapien
8	142.5	26.1	544	4	Q9BZG7	Q9BZG7 homo sapien
9	141	25.9	221	10	O65514	O65514 arabidopsis
10	141	25.9	318	10	Q38777	Q38777 allium sati
11	139.5	25.6	233	16	Q92NU7	Q92NU7 rhizobium m
12	139.5	25.6	393	5	Q18880	Q18880 caenorhabdi
13	139	25.5	335	10	O65330	O65330 elaeagnus u
14	138	25.3	207	10	Q43522	Q43522 lycopersico
15	137	25.1	697	5	Q9GRW7	Q9GRW7 drosophila
16	136.5	25.0	718	12	Q91TR1	Q91TR1 tupala herp

17	135	24.8	163	5	Q95UX4	Q95ux4 drosophila
18	135	24.8	165	5	Q9GP44	Q9gp44 drosophila
19	135	24.8	165	5	Q95NR6	Q95nr6 drosophila
20	134.5	24.7	165	5	Q95UX2	Q95ux2 drosophila
21	134.5	24.7	192	16	Q92P87	Q92p87 rhizobium m
22	134.5	24.7	447	13	Q73628	Q73628 anolis caro
23	134.5	24.7	452	10	Q9LW52	Q9lw52 arabidopsis
24	134.5	24.7	1084	10	Q43010	Q43010 oryza sativ
25	133.5	24.5	441	11	Q99MY1	Q99my1 mus musculu
26	133	24.4	106	5	Q9GP80	Q9gp80 drosophila
27	132	24.2	132	10	Q943G4	Q943g4 oryza sativ
28	131.5	24.1	163	5	Q95UW9	Q95uw9 drosophila
29	130.5	23.9	162	5	Q95UX5	Q95ux5 drosophila
30	130.5	23.9	164	5	Q95NP2	Q95np2 drosophila
31	130.5	23.9	175	10	Q9LSN6	Q9lsn6 arabidopsis
32	130	23.9	159	5	Q95UW1	Q95uw1 drosophila
33	130	23.9	161	5	Q95UX3	Q95ux3 drosophila
34	130	23.9	163	5	Q95NU6	Q95nu6 drosophila
35	130	23.9	168	5	Q95UW8	Q95uw8 drosophila
36	130	23.9	255	10	Q9SIH2	Q9sih2 arabidopsis
37	130	23.9	291	10	Q39337	Q39337 brassica na
38	129.5	23.8	738	5	O02402	O02402 pinctada fu
39	129	23.7	106	10	Q9SHA4	Q9sha4 arabidopsis
40	129	23.7	155	5	Q9GP74	Q9gp74 drosophila
41	129	23.7	155	5	Q9GND8	Q9gnd8 drosophila
42	129	23.7	321	6	Q9MYX6	Q9myx6 ovis aries
43	129	23.7	322	4	Q96EW0	Q96ew0 homo sapien
44	129	23.7	337	11	Q9ET75	Q9et75 rattus norv
45	129	23.7	698	5	Q9GRX4	Q9grx4 drosophila

ALIGNMENTS

RESULT 1	
O64033	
ID O64033	PRELIMINARY; PRT; 56 AA.
AC O64033;	
DT 01-AUG-1998	(TREMBLrel. 07, Created)
DT 01-AUG-1998	(TREMBLrel. 07, Last sequence update)
DT 01-DEC-2001	(TREMBLrel. 19, Last annotation update)
DE	PUTATIVE LIPOPROTEIN.

```

Query Match          55.8%; Score 304; DB 9; Length 56;
Best Local Similarity 100.0%; Pred. No. 5.8e-26;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MEKLFKEVKLEELLENQKSGSLGKAQCAALWLCASGGTICGGGAVACONTYQFCR 56
|||||
1 MEKLFKEVKLEELLENQKSGSLGKAQCAALWLCASGGTICGGGAVACONTYQFCR 56

RESULT 2	
O34781	
ID O34781	PRELIMINARY; PRT; 56 AA.
AC O34781;	
DT 01-JAN-1998	(TREMBLrel. 05, Created)
DT 01-JAN-1998	(TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998	(TREMBLrel. 08, Last annotation update)


```
Query Match      27.0%; Score 147; DB 10; Length 333;
Best Local Similarity 54.5%; Pred. No. 4.3e-08;
Matches 36; Conservative 2; Mismatches 16; Indels 12; Gaps 4;

QY 16 QGSGGLGKAQCAALW-LQCASGGTIG- - - - -CGGAVACQNYRQFCRGGGGGGGGGGGG 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25 QCGSQAGGALCPNC-LCCSSYGGWGSTDYCGDG- - - - -CQSQ- - - - -CDGCGGGGGGGGGGG 77
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 71 GGGGGG 76
   |||||
Db 78 GGGGGG 83
   |||||

RESULT 5
Q9BZG5 PRELIMINARY; PRT; 531 AA.
AC Q9BZG5;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ANDROGEN RECEPTOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Magklara A., Diamandis E.P.;
RT "Poly-Q and poly-G repeats in the androgen receptor of the breast
   cancer cell line MCF-7."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF321917; AAK09426.1; -
DR InterPro; IPR001103; Androgen_recep.
DR Pfam; PF02166; Androgen_recep; 1.
KW Receptor.
FT NON_TER 531
SQ SEQUENCE 531 AA; 54419 MW; FA0D3AA60AC20DAB CRC64;

Query Match      26.1%; Score 142.5; DB 4; Length 531;
Best Local Similarity 50.8%; Pred. No. 2.2e-07;
Matches 32; Conservative 3; Mismatches 12; Indels 16; Gaps 3;

QY 18 GSGGLGKAQCAALW- - - - -LQCASGGTI- - - - -GCGGAVACQNYRQFCRGGGGGGGGGG 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 415 GSGSPSAASSSSWHTLFTAEQGQLYGPGCGG- - - - -GGGGGGGGGGGGGGGGGG 462
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 74 GGG 76
   |||
Db 463 GGG 465
   |||

RESULT 6
Q9NUA2 PRELIMINARY; PRT; 539 AA.
AC Q9NUA2;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE DJ80804.1 (ANDROGEN RECEPTOR (DIHYDROTESTOSTERONE RECEPTOR))
DE (FRAGMENT).
GN AR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chapman J.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Magklara A., Diamandis E.P.;
RT "Poly-Q and poly-G repeats in the androgen receptor of the breast
```

```
RT cancer cell line T-47D."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049564; CAB87955.1; -
DR EMBL; AF321915; AAK09424.1; -
DR InterPro; IPR001103; Androgen_recep.
DR Pfam; PF02166; Androgen_recep; 1.
KW Receptor.
FT NON_TER 539
SQ SEQUENCE 539 AA; 55444 MW; AB493953B89D869F CRC64;

Query Match      26.1%; Score 142.5; DB 4; Length 539;
Best Local Similarity 50.8%; Pred. No. 2.2e-07;
Matches 32; Conservative 3; Mismatches 12; Indels 16; Gaps 3;

QY 18 GSGGLGKAQCAALW- - - - -LQCASGGTI- - - - -GCGGAVACQNYRQFCRGGGGGGGGGGGG 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 423 GSGSPSAASSSSWHTLFTAEQGQLYGPGCGG- - - - -GGGGGGGGGGGGGGGGGG 470
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 74 GGG 76
   |||
Db 471 GGG 473
   |||

RESULT 7
Q9BZG6 PRELIMINARY; PRT; 542 AA.
AC Q9BZG6;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE ANDROGEN RECEPTOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Magklara A., Diamandis E.P.;
RT "Poly-Q and poly-G repeats in the androgen receptor of the breast
   cancer cell line BT-474."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF321916; AAK09425.1; -
DR InterPro; IPR001103; Androgen_recep.
DR Pfam; PF02166; Androgen_recep; 1.
KW Receptor.
FT NON_TER 542
SQ SEQUENCE 542 AA; 55828 MW; C363EF841CAF7739 CRC64;

Query Match      26.1%; Score 142.5; DB 4; Length 542;
Best Local Similarity 50.8%; Pred. No. 2.2e-07;
Matches 32; Conservative 3; Mismatches 12; Indels 16; Gaps 3;

QY 18 GSGGLGKAQCAALW- - - - -LQCASGGTI- - - - -GCGGAVACQNYRQFCRGGGGGGGGGGGG 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 426 GSGSPSAASSSSWHTLFTAEQGQLYGPGCGG- - - - -GGGGGGGGGGGGGGGGGG 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 74 GGG 76
   |||
Db 474 GGG 476
   |||

RESULT 8
Q9BZG7 PRELIMINARY; PRT; 544 AA.
AC Q9BZG7;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE ANDROGEN RECEPTOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```

```
RN
RP SEQUENCE FROM N.A.
RA Magklara A., Diamandis E.P.;
RT "Poly-Q and poly-G repeats in the androgen receptor of the breast
RL cancer cell line ZR-75-1."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF321914; AAK09423.1; -
DR InterPro; IPR001103; Androgen_recep.
DR Pfam; PF02166; Androgen_recep; 1.
KW Receptor.
FT NON_TER 544 544
SQ SEQUENCE 544 AA; 56013 MW; C97133EB6C922E74 CRC64;

Query Match 26.1%; Score 142.5; DB 4; Length 544;
Best Local Similarity 50.8%; Pred. No. 2.2e-07;
Matches 32; Conservative 3; Mismatches 13; Indels 15; Gaps 3;

QY 18 GSGLGKAGCAALM--LQASGGTI--GCGGGAVACQNYRQFCRGGGGGGGGGGGGG 73
Db 427 GSGSPSAASSSWHTLFTAEQGQLYGPCGGG-----GGGGGGGGGGGGGGG 475

QY 74 GGG 76
Db 476 GGG 478

RESULT 9
O65514 PRELIMINARY; PRT; 221 AA.
AC O65514;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE PUTATIVE GLYCINE-RICH CELL WALL PROTEIN.
GN F23E13.120 OR AT4G36230.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Hoheisel J., Jesse T., Heijnen L., Vos P., Mewes H.W., Mayer K.,
RA Schueller C.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022141; CAA18129.1; -.
DR EMBL; AL161589; CAB80294.1; -.
SQ SEQUENCE 221 AA; 20344 MW; DE8DAE2E2C57A9F8 CRC64;

Query Match 25.9%; Score 141; DB 10; Length 221;
Best Local Similarity 65.1%; Pred. No. 1.2e-07;
Matches 28; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 34 ASGGTICGGGAVACQNYRQFCRGGGGGGGGGGGGGGGGGGGGG 76
Db 130 SGGGGGGGGGGGSGNGSGRCRGGGGGGGGGGGGGGGGGGG 172

RESULT 10
Q38777 PRELIMINARY; PRT; 318 AA.
F9 Q38777
AC Q38777;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
```

```
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CHITINASE PRECURSOR (FRAGMENT).
GN CHITINASE.
OS Allium sativum (Garlic).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;
OC Allium.
OX NCBI_TaxID=4682;
RN [1]
RP SEQUENCE FROM N.A.
RA van Damme E.J.M., Willems P., Peumans W.;
RT "Isolation and characterization of two different chitinase cDNA clones
RL from garlic (Allium sativum L.) shoots."
RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M94105; AAA32641.1; -.
DR HSP; P23951; 2BAA.
DR InterPro; IPR001002; Chitin_bind.
DR InterPro; IPR000726; Glyco_hydro_19.
DR Pfam; PF00187; Chitin_binding; 1.
DR Pfam; PF00182; Glyco_hydro_19; 1.
DR PRINTS; PR00451; CHITINBINDNG.
DR ProDom; PD000574; Glyco_hydro_19; 1.
DR ProDom; PD000609; Chitin_bind; 1.
DR SMART; SM00270; ChtBD1; 1.
DR PROSITE; PS00773; CHITINASE_19_1; 1.
DR PROSITE; PS00774; CHITINASE_19_2; 1.
DR PROSITE; PS00026; CHITIN_BINDING; 1.
KW Chitin-binding; Signal.
FT NON_TER 1 1
FT SIGNAL <1 17 POTENTIAL.
FT CHAIN 18 318 CHITINASE.
SQ SEQUENCE 318 AA; 33175 MW; 097C3DBD772468D4 CRC64;

Query Match 25.9%; Score 141; DB 10; Length 318;
Best Local Similarity 46.2%; Pred. No. 1.8e-07;
Matches 37; Conservative 5; Mismatches 22; Indels 16; Gaps 5;

QY 4 LFEVKLELEENQKSGSLGKAGCAALMQLQASGGTIG-----CGGGAVACQNYRQFCRGG 58
Db 11 LFK----NSYAQQCGSQAGGALCSNR-LCCSKFGYCGSFDPCGTG---CQSQ---CGGG 59

QY 59 GGGGGGGGGGGGGGGGGGGGMS 78
Db 60 GGGGGGGGGGGGGGGGGGVA 79

RESULT 11
Q92NU7 PRELIMINARY; PRT; 233 AA.
AC Q92NU7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE GLYCINE-RICH PROTEIN.
GN SMC04357.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21368234; PubMed=11474104;
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar A., Hyman R.W., Jones T., Kahn D.,
RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
RA Ramsperger U., Surzycki R., Thebault P., Vandenbol M.,
```


ID Q9GRW7 PRELIMINARY; PRT; 697 AA.
AC Q9GRW7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NONA PROTEIN.
GN NONA.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7244;
RN [1]
RP SEQUENCE FROM N.A.
RA Campesan S., Chalmers D., Sandrelli F., Megighian A., Peixoto A.A.,
RA Costa R., Kyriacou C.P.;
RT "Comparative analysis of the nonA region in Drosophila identifies
RT highly diverged 5' gene that may constrain nonA promoter evolution.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ298998; CAC10058.1; -.
DR HSSP; P11940; ICVJ.
DR FlyBase; FBgn0042729; Dvir\nonA.
DR InterPro; IPR000504; RRM.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS0102; RRM; 2.
SQ SEQUENCE 697 AA; 75337 MW; 516BD031DA5FABFB CRC64;

Query Match 25.1%; Score 137; DB 5; Length 697;
Best Local Similarity 57.1%; Pred. No. 1.1e-06;
Matches 28; Conservative 1; Mismatches 8; Indels 12; Gaps 1;

QY 34 ASGGTIGCGGAVACQNYRQFCRGGGGGGGGGGGGGGGGGGGGGMSKFDD 82
| | | | |
Db 194 ARGGGGGGGG-----GGGGGGGGGGGGGGGGGGGGGRDRNPD 230

Search completed: October 24, 2002, 09:29:12
Job time : 35.1168 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:23:41 ; Search time 14.3139 Seconds
(without alignments)
287.115 Million cell updates/sec

Title: US-09-894-030-3_COPY_1_37
Perfect score: 193
Sequence: 1 MEKLFKEVLEKLENGKSGGLGKAQCAALWLQASGG 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	193	100.0	56	20 AAY03205	Amino acid sequenc
2	100	51.8	37	20 AAY03207	Amino acid sequenc
3	96	49.7	37	20 AAY03218	Amino acid sequenc
4	58	30.1	408	21 AAG16276	Arabidopsis thalia
5	58	30.1	464	21 AAG16275	Arabidopsis thalia
6	58	30.1	464	22 AAB47604	Cystathionine beta
7	58	30.1	464	22 AAB20017	Arabidopsis cystat
8	58	30.1	506	21 AAG16274	Arabidopsis thalia
9	56	29.0	200	21 AAG58269	Arabidopsis thalia
10	56	29.0	200	21 AAG60923	Arabidopsis thalia
11	56	29.0	202	21 AAG58268	Arabidopsis thalia

12	56	29.0	202	21	AAG60922	Arabidopsis thalia
13	56	29.0	215	21	AAG60921	Arabidopsis thalia
14	56	29.0	216	21	AAG58267	Arabidopsis thalia
15	55.5	28.8	70	20	AAY24125	Conopeptide propep
16	55.5	28.8	84	22	AAU05933	Cone snail O-supe
17	55	28.5	103	21	AAG02572	Human secreted pro
18	55	28.5	740	22	ABG28146	Novel human diago
19	54.5	28.2	839	20	AAY30152	A human vanilloid
20	54.5	28.2	839	20	AAY30153	A partial human va
21	54.5	28.2	839	20	AAY30155	A human vanilloid
22	54.5	28.2	839	20	AAY06558	Human capsaicin re
23	54.5	28.2	839	21	AAB32127	Human vanilloid re
24	54.5	28.2	839	21	AAY96478	Human vanilloid re
25	54.5	28.2	839	21	AAY97357	Human VR-1 protein
26	54.5	28.2	839	22	AAE01229	Human vanilloid re
27	53.5	27.7	82	22	AAU05937	Cone snail O-supe
28	53	27.5	907	22	ABB61537	Drosophila melanog
29	53	27.5	297	18	AAW20066	Thielavia terrestr
30	53	27.5	467	20	AAW81021	Fragment of cystat
31	53	27.5	905	22	ABG29396	Novel human diago
32	52	26.9	223	20	AAW81019	Fragment of cystat
33	51.5	26.7	289	22	AAB64964	Gene 24 human secr
34	51.5	26.7	378	22	AAB64962	Novel human diago
35	51.5	26.7	644	22	ABG18134	H. pylori derived
36	50.5	26.2	63	18	AAW20382	Human transcriptio
37	50.5	26.2	765	22	ABB50196	Human polypeptide
38	50.5	26.2	786	22	AAM41413	Human protein sequ
39	50.5	26.2	842	22	AAB95612	Human polypeptide
40	50.5	26.2	898	22	AAM39627	Novel human secret
41	50	25.9	221	22	AAU30375	Human polypeptide
42	50	25.9	374	22	AAM39000	Human polypeptide
43	50	25.9	375	20	AAY32885	Soybean flavanone-
44	50	25.9	481	22	AAM40786	Human polypeptide
45	50	25.9	482	21	AAB23631	Human secreted pro

ALIGNMENTS

RESULT 1
AAY03205
ID AAY03205 standard; Protein; 56 AA.
XX AC AAY03205;
XX DT 03-AUG-1999 (first entry)
XX DE Amino acid sequence of sunA protein.
XX KW Sublancin 168; antimicrobial; food preservative; pro-sublancin 168;
XX KW Gram-positive bacteria; pre-sublancin 168; sunA; sunT.
XX OS Bacillus subtilis.
XX PN WO9903352-A1.
XX PD 28-JAN-1999.
XX PF 17-JUL-1998; 98WO-US14547.
XX PR 18-JUL-1997; 97US-0053035.
XX (UYMA-) UNIV MARYLAND BALTIMORE.
XX PI Hansen JN;
XX DR WPI; 1999-131752/11.
XX DR N-PSDB; AAX28631.
XX PT New antimicrobial peptide, sublancin 168, from Bacillus subtilis -
XX PT used for, e.g. treatment of infections caused by Gram negative
XX PT bacteria and as food preservative

PS Claim 6; Fig 3; 71pp; English.

XX This is the amino acid sequence of the suna protein encoded by the
 CC prosublancin 168 nucleotide sequence used in the method of the
 CC invention. The peptide designated sublancin 168, is an antimicrobial
 CC useful for treating infections and preserving food against spoilage
 CC bacteria, particularly Gram-positive bacteria. Pro-sublancin 168 and
 CC pre-sublancin 168, are the precursors of sublancin 168. Sublancin
 CC 168 is very stable at low pH and can be autoclaved without damage.
 CC It does not decompose after 2 years in aqueous solution of about
 CC neutral pH.

XX Sequence 56 AA;

SQ Query Match 100.0%; Score 193; DB 20; Length 56;
 Best Local Similarity 100.0%; Pred. No. 1.9e-18;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKLFKEVKELEENQKSGGLGKAQCAALWLQASGG 37
 Db 1 MEKLFKEVKELEENQKSGGLGKAQCAALWLQASGG 37

RESULT 2

AAAY03207
 ID AAY03207 standard; Protein; 37 AA.
 XX AAY03207;
 AC AAY03207;
 DT 03-AUG-1999 (first entry)
 DE Amino acid sequence of prosublancin 168.
 XX Sublancin 168; antimicrobial; food preservative; pro-sublancin 168;
 KW Gram-positive bacteria; pre-sublancin 168.
 XX Bacillus subtilis.
 OS WO9903352-A1.
 PN 28-JAN-1999.
 XX 17-JUL-1998; 98WO-US14547.
 PF 18-JUL-1997; 97US-0053035.
 XX (UYMA-) UNIV MARYLAND BALTIMORE.
 PA Hansen JN;
 PI WPI; 1999-131752/11.
 DR New antimicrobial peptide, sublancin 168, from Bacillus subtilis -
 XX used for, e.g. treatment of infections caused by Gram negative
 PT bacteria and as food preservative
 PT Claim 1; Pages 52-53; 71pp; English.
 XX This is the amino acid sequence of prosublancin 168 used in the
 CC method of the invention. The peptide designated sublancin 168, is
 CC an antimicrobial useful for treating infections and preserving food
 CC against spoilage bacteria, particularly Gram-positive bacteria.
 CC Pro-sublancin 168 and pre-sublancin 168, are the precursors of
 CC sublancin 168. Sublancin 168 is very stable at low pH and can be
 CC autoclaved without damage. It does not decompose after 2 years in
 CC aqueous solution of about neutral pH.

XX Sequence 37 AA;

SQ Query Match 51.8%; Score 100; DB 20; Length 37;
 Best Local Similarity 100.0%; Pred. No. 3.1e-06;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GLGKAQCAALWLQASGG 37
 Db 1 GLGKAQCAALWLQASGG 18

RESULT 3

AAAY03218
 ID AAY03218 standard; Protein; 37 AA.
 XX AAY03218;
 AC AAY03218;
 DT 03-AUG-1999 (first entry)
 DE Amino acid sequence of sublancin 168.
 XX Sublancin 168; antimicrobial; food preservative; pro-sublancin 168;
 KW Gram-positive bacteria; Pro-sublancin 168; pre-sublancin 168.
 XX Bacillus subtilis.
 OS Key Location/Qualifiers
 FH Disulfide-bond 7..36
 FT Disulfide-bond 14..29
 FT Modified-site 16
 FT /note= "dehydrogenated Ser (Dha)"
 FT Modified-site 19
 FT /note= "dehydrogenated Thr (Dhb)"
 FT Cross-links 19..22
 FT /note= "thioether bridge"
 FT
 XX WO9903352-A1.
 PN 28-JAN-1999.
 PD 17-JUL-1998; 98WO-US14547.
 PF 18-JUL-1997; 97US-0053035.
 XX (UYMA-) UNIV MARYLAND BALTIMORE.
 PA Hansen JN;
 PI WPI; 1999-131752/11.
 DR New antimicrobial peptide, sublancin 168, from Bacillus subtilis -
 XX used for, e.g. treatment of infections caused by Gram negative
 PT bacteria and as food preservative
 PT Claim 13; Fig 7; 71pp; English.
 XX This is the amino acid sequence of sublancin 168 used in the
 CC method of the invention. The peptide designated sublancin 168, is
 CC an antimicrobial useful for treating infections and preserving food
 CC against spoilage bacteria, particularly Gram-positive bacteria.
 CC Pro-sublancin 168 and pre-sublancin 168, are the precursors of
 CC sublancin 168. Sublancin 168 is very stable at low pH and can be
 CC autoclaved without damage. It does not decompose after 2 years in
 CC aqueous solution of about neutral pH.

XX Sequence 37 AA;

SQ Query Match 49.7%; Score 96; DB 20; Length 37;
 Best Local Similarity 94.4%; Pred. No. 1e-05;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 20 GLGKAQCAALWLQASGG 37
 Db 1 GLGKAQCAALWLQASGG 18

RESULT 4

AAG16276
 ID AAG16276 standard; Protein; 408 AA.

XX AAG16276; 99US-0139462.
AC 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
DT 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
DE 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
KW 23-JUN-1999; 99US-0140354.
KW 24-JUN-1999; 99US-0140695.
KW 28-JUN-1999; 99US-0140823.
XX 29-JUN-1999; 99US-0140991.
OS 30-JUN-1999; 99US-0141287.
XX 01-JUL-1999; 99US-0141842.
PN 01-JUL-1999; 99US-0142154.
XX 02-JUL-1999; 99US-0142055.
PD 06-JUL-1999; 99US-0142390.
XX 08-JUL-1999; 99US-0142803.
PF 09-JUL-1999; 99US-0142920.
XX 12-JUL-1999; 99US-0142977.
XX 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.

XX Arabidopsis thaliana protein fragment SEQ ID NO: 16857.
AC Protein identification; signal transduction pathway; metabolic pathway;
DT hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
DE Arabidopsis thaliana.
XX EP1033405-A2.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
PF 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.

PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 30.1%; Score 58; DB 21; Length 408;
Best Local Similarity 48.6%; Pred. No. 15;
Matches 17; Conservative 2; Mismatches 10; Indels 6; Gaps 3;

QY 2 EKLFEVKLEELNKGSGLGKQAALWLQCAAG 36
||| ||| | : | : ||| | | | | |
Db 240 EKLAKVYF--LQNSEGSLAPFDC--WL-CLRG 268

RESULT 5
AAG16275
ID AAG16275 standard; Protein; 464 AA.

XX AC AAG16275;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 16856.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 30.1%; Score 58; DB 21; Length 464;
Best Local Similarity 48.6%; Pred. No. 18;
Matches 17; Conservative 2; Mismatches 10; Indels 6; Gaps 3;

Qy 2 EKLFKEVKLEELEENQKSGGLGKAQCAALWLQCASG 36
||| ||| |:|:|||| | || |
Db 296 EKLAKEVYF--LQNSEGSGLAPFDC---WL-CLRG 324

RESULT 6
AAB47604
ID AAB47604 standard; Protein; 464 AA.
XX
AC AAB47604;
XX
DT 07-JAN-2002 (first entry)
XX
DE Cystathionine beta lyase, CBL.
XX Herbicide; porphobilinogen deaminase; crop; CBL; ENR-A; UROD;
KW enoyl-acyl carrier protein reductase; cystathionine beta lyase;
KW uroporphyrinogen decarboxylase; PBGD; porphobilinogen deaminase;
KW CPPO; coproporphyrinogen oxidase.
XX
OS Arabidopsis thaliana.
XX

PN US6294345-B1.
XX PD 25-SEP-2001.
XX PF 05-JUN-2000; 2000US-0586719.
XX PR 27-JUL-1999; 99US-198218P.
XX PR 20-AUG-1999; 99US-240929P.
XX PR 22-NOV-1999; 99US-228810P.
XX PR 27-MAR-2000; 2000US-287572P.
XX PR 22-MAY-2000; 2000US-287571P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX Levin JZ, Bauer MW, Zheng F;
XX WPI; 2001-638012/73.
XX N-PSDB; AAH43576.
XX
PT Identifying herbicidal compounds for suppressing growth of undesirable
PT vegetation in fields where important crops are grown, comprises
PT combining a plant PBGD polypeptide with a test compound -
XX
XX Example 21; Column 55-58; 47pp; English.
XX
XX The sequences given in AAB47603-07 are proteins which may be used
XX to identify herbicidal compounds, by combining one of these plant
XX polypeptides, which have porphobilinogen deaminase activity, with a
XX test compound under conditions allowing binding or inhibition.
XX The method may be used for identifying herbicidal compounds, which can
XX be used to suppress the growth of undesirable vegetation in fields where
XX important crops are grown. The cDNA sequences encoding these proteins
XX may be used to confer herbicide tolerance to plants, plant cells or
XX tissues.
XX
SQ Sequence 464 AA;
Query Match 30.1%; Score 58; DB 22; Length 464;
Best Local Similarity 48.6%; Pred. No. 18;
Matches 17; Conservative 2; Mismatches 10; Indels 6; Gaps 3;
QY 2 EKLFKEVKLEELEENQKSGGLGKAQCAALWLCASG 36
||| ||| | :| :||| | || | |
Db 296 EKLAKEVYF--LQNSEGSLAPFDC---WL-CLRG 324
RESULT 7
AAB20017
ID AAB20017 standard; Protein; 464 AA.
XX AAB20017;
XX
XX 28-MAR-2001 (first entry)
XX
XX Arabidopsis cystathionine beta lyase.
XX Cystathionine beta lyase; CBL; herbicide tolerance;
XX transgenic plant.
XX Arabidopsis thaliana.
XX
XX WO200077185-A2.
XX
XX 21-DEC-2000.
XX
XX 13-JUN-2000; 2000WO-EP05432.
XX
XX 15-JUN-1999; 99US-0333366.
XX 27-JUL-1999; 99US-0361879.
XX 20-AUG-1999; 99US-0378313.
XX 22-NOV-1999; 99US-0444117.
XX 01-DEC-1999; 99US-0452671.
XX

PA (NOVS) NOVARTIS AG.
XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX Levin JZ, Bauer MW, Zheng F;
XX WPI; 2001-041382/05.
XX N-PSDB; AAA89292.
XX
PT New Arabidopsis thaliana DNA molecules for identifying compounds that
PT are inhibitors of Arabidopsis enzymes encoded by the DNA, which can be
PT used as herbicides -
XX
XX Disclosure; Page 74-75; 93pp; English.
XX
XX The present sequence is that of Arabidopsis thaliana
XX cystathionine beta lyase (CBL, EC-4.4.1.8), an enzyme that
XX catalyses the conversion of cystathionine to homocysteine. The
XX enzyme is essential for normal plant growth and development. The
XX invention provides ENR-A, CBL, UROD, PBGD and CPPO genes (see
XX AAA89291-95), methods for the recombinant production of the encoded
XX enzymes (see AAB20016-20) in heterologous hosts, methods for
XX screening chemicals for herbicidal activity using these recombinant
XX enzymes, and methods for using these herbicides to suppress growth
XX of undesired vegetation. Plants, plant tissue, seeds and cells
XX tolerant to herbicides that inhibit ENR-A, CBL, UROD, PBGD or CPPO
XX are obtained by altering the corresponding enzyme activity either by
XX increasing expression of wild-type herbicide-sensitive enzymes or
XX by expressing modified herbicide tolerant enzymes. The herbicide
XX tolerant enzymes can also be used as selectable markers.
XX
SQ Sequence 464 AA;
Query Match 30.1%; Score 58; DB 22; Length 464;
Best Local Similarity 48.6%; Pred. No. 18;
Matches 17; Conservative 2; Mismatches 10; Indels 6; Gaps 3;
QY 2 EKLFKEVKLEELEENQKSGGLGKAQCAALWLCASG 36
||| ||| | :| :||| | || | |
Db 296 EKLAKEVYF--LQNSEGSLAPFDC---WL-CLRG 324
RESULT 8
AAG16274
ID AAG16274 standard; Protein; 506 AA.
XX AAG16274;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 16855.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX

PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 30.1%; Score 58; DB 21; Length 506;
Best Local Similarity 48.6%; Pred. No. 19;
Matches 17; Conservative 2; Mismatches 10; Indels 6; Gaps 3;

QY 2 EKLPKEVKLEENQKSGGLGKAQCAALWLQACSG 36
||| ||| | : | : ||| | || |
Db 338 EKLAKVYF--LQNSEGSLAPFDC---WL-CLRG 366

RESULT 9
AAG58269
ID AAG58269 standard; Protein; 200 AA.

AC AAG58269;
XX
XX
DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 75196.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.
XX
XX EP1033405-A2.

XX
PD 06-SEP-2000.

XX
PF 25-FEB-2000; 2000EP-0301439.

XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.

RESULT 10	
AAG60923	
ID	AAG60923 standard; Protein; 200 AA.
XX	
XX	
AC	AAG60923;
XX	
DT	18-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 78966.
XX	
KW	Protein identification; signal transduction pathway; metabolic pathway;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	termination sequence.
XX	
OS	Arabidopsis thaliana.
XX	
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 29.0%; Score 56; DB 21; Length 200;
Best Local Similarity 43.3%; Pred. No. 13;
Matches 13; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

QY 4 LFEKVELEENQKSGGLGKAQCAAL--WL 31
||||: ||| |::||| |||: ||
Db 128 LFKKLDVEGLSEIGAIGLAAMGCAAMFAWL 157

RESULT 11

AAG58268
ID AAG58268 standard; Protein; 202 AA.

XX
AC AAG58268;

XX
DT 18-OCT-2000 (first entry)

XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 75195.

XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX
OS Arabidopsis thaliana.

XX
PN EP1033405-A2.

XX
PD 06-SEP-2000.

XX
PF 25-FEB-2000; 2000EP-0301439.

XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 22-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.

PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.
 Query Match 29.0%; Score 56; DB 21; Length 202;
 Best Local Similarity 43.3%; Pred. No. 13;
 Matches 13; Conservative 5; Mismatches 10; Indels 2; Gaps 1;
 QY 4 LFKVKKLEELNKGSGGLGKAQCAAL--WL 31
 |||::||| |::| |::| |
 Db 130 LFKKLDVEGLSEAGAGLAAMGCAAMFAWL 159
 RESULT 12
 AAG60922
 ID AAG60922 standard; Protein; 202 AA.
 XX
 AC AAG60922;
 XX 18-OCT-2000 (first entry)
 DT Arabidopsis thaliana protein fragment SEQ ID NO: 78965.
 DE Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX Arabidopsis thaliana.
 OS EP1033405-A2.
 XX 06-SEP-2000.
 PD 25-FEB-2000; 2000EP-0301439.
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

29.0%; Score 56; DB 21; Length 202;

Best Local Similarity 43.3%; Pred. No. 13;

Matches 13; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

QY 4 LFKEVKLEELNKGSGLGKAOCAAL--WL 31
|||: :| | | :||| |||: ||
Db 130 LFKKLDVEGLSEAIGAGLAAMGCAAMFAWL 159

RESULT 13

AAG60921

ID AAG60921 standard; Protein; 215 AA.

XX AC AAG60921;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 78964.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149930.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 29.0%; Score 56; DB 21; Length 215;
 Best Local Similarity 43.3%; Pred. No. 14;
 Matches 13; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

QY 4 LFKVKKLEENKSGSLGKAQCAAL--WL 31
 |||:::| | | | | | | | | | | |
 Db 143 LFKKLDVEGLSEAGLAAGLAAMGCAAMFAWL 172

RESULT 14

AAG58267
 ID AAG58267 standard; Protein; 216 AA.
 XX
 AC AAG58267;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 75194.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.

PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	24-SEP-1999;	99US-0155486.
PR	28-SEP-1999;	99US-0155659.
PR	29-SEP-1999;	99US-0156458.
PR	04-OCT-1999;	99US-0156596.
PR	05-OCT-1999;	99US-0157117.
PR	06-OCT-1999;	99US-0157753.
PR	07-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 29.0%; Score 56;		
Best Local Similarity 43.3%; Pred. No.		
Matches 13; Conservative 5; Mismatch		
QY	4 LFKVKLEELNKGSGLGKAQCAAL--WL 31	:::
Db	144 LFKKLDVEGLSEATGAGLAAMGCAAMFAWL 173	:::
RESULT 15		
AAY24125		
ID	AAY24125 standard; peptide; 70 AA.	
XX	AAY24125;	
AC	AC	
XX	XX	
DT	13-SEP-1999 (first entry)	
XX	XX	

DE Conopeptide propeptide Mr6.3.
XX
KW Conopeptide; gamma-conopeptide; venom; cone snail; cation channel;
KW epilepsy; pacemaker; heart muscle; neuronal pacemaker calcium channel.
XX
OS Conus marmoreus.
XX
PN W09930732-A1.
XX
PD 24-JUN-1999.
XX
PF 16-DEC-1998; 98WO-US26792.
XX
PR 16-DEC-1997; 97US-0069706.
XX
PA (REGC) UNIV CALIFORNIA.
PA (UTAH) UNIV UTAH RES FOUND.
PA (UYVR-) UNIV VRIJE.
XX
PI Burlingame AL, Colledge C, Cruz LJ, Fainzilber M;
PI Imperial J, Kits KS, Olivera BM, Shetty R, Walker C;
PI Walkins M;
XX
DR WPI; 1999-418708/35.
DR N-PSDB; AAX88505.
XX
PT Gamma-carboxylated conopeptides used as, e.g. neuronal pacemaker
PT calcium channels
XX
PS Claim 37; Page 52; 61pp; English.
XX
CC The present invention describes gamma-carboxylated conopeptides derived
CC from cone snail venom. The gamma-conopeptides and their propeptides are
CC useful as agonists of neuronal pacemaker calcium channels. The
CC conopeptides are naturally available in minute amounts in the venom of
CC cone snails and their derivatives are synthetic. The peptides modulate
CC slow inward cation channels in vertebrates involved in syndromes of
CC clinical relevance, such as epileptic activity in hippocampus and
CC pacemaker potentials in heart muscle. The present sequence represents
CC a conopeptide propeptide.
XX
SQ Sequence 70 AA;

Query Match 28.8%; Score 55.5; DB 20; Length 70;
Best Local Similarity 30.0%; Pred. NO. 5.1;
Matches 12; Conservative 9; Mismatches 14; Indels 5; Gaps 1;

QY 2 EKLFKEVKLELENQKGGGLGKAQ-----CAALWLQCASG 36
Db 18 QALYQEKRRKEMINFLSKGINAERRNEGCKATWMSGSSG 57

Search completed: October 24, 2002, 09:28:00
Job time : 16.3139 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:27:01 ; Search time 5.67153 Seconds
(without alignments)
159.348 Million cell updates/sec

Title: US-09-894-030-3_COPY_1_37
Perfect score: 193
Sequence: 1 MEKLFKEVKLELENQKSGGLGKAQCAALWLQCASGG 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA: *
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	30.1	464	US-09-586-719-4	Sequence 4, Appli
2	54.5	28.2	839	US-09-197-636-2	Sequence 2, Appli
3	54.5	28.2	839	US-09-197-636-4	Sequence 4, Appli
4	54.5	28.2	839	US-09-197-636-8	Sequence 8, Appli
5	54.5	28.2	839	US-09-235-451-34	Sequence 34, Appl
6	51	26.4	549	US-08-325-071-61	Sequence 61, Appl
7	51	26.4	549	US-08-461-004A-61	Sequence 61, Appl
8	51	26.4	620	US-08-325-071-65	Sequence 65, Appl
9	51	26.4	620	US-08-461-004A-65	Sequence 65, Appl
10	51	26.4	650	US-08-325-071-63	Sequence 63, Appl
11	51	26.4	650	US-08-325-071-67	Sequence 67, Appl
12	51	26.4	650	US-08-461-004A-63	Sequence 63, Appl
13	51	26.4	650	US-08-461-004A-67	Sequence 67, Appl
14	49	25.4	259	US-08-402-804-8	Sequence 8, Appli
15	49	25.4	1296	US-08-470-260-3	Sequence 3, Appli
16	49	25.4	1296	US-08-471-491-3	Sequence 3, Appli
17	49	25.4	1296	US-08-466-662-3	Sequence 3, Appli
18	48.5	25.1	1287	US-08-200-232-2	Sequence 2, Appli
19	48.5	25.1	1287	PCT-US95-02219-2	Sequence 2, Appli
20	48.5	25.1	1287	PCT-US95-02219A-2	Sequence 2, Appli
21	48	24.9	581	US-08-724-394A-3	Sequence 3, Appli
22	47	24.4	428	US-09-118-319-5	Sequence 5, Appli
23	47	24.4	500	US-09-141-000-2	Sequence 2, Appli
24	47	24.4	1040	US-08-254-989-2	Sequence 2, Appli
25	46.5	24.1	728	US-08-981-392-2	Sequence 2, Appli
26	46.5	24.1	729	US-08-872-855-8	Sequence 8, Appli
27	46	23.8	215	US-09-220-528-104	Sequence 104, App

28	46	23.8	284	5	PCT-US94-09752-4	Sequence 4, Appli
29	46	23.8	591	1	US-08-145-995A-21	Sequence 21, Appl
30	46	23.8	591	2	US-08-451-747-21	Sequence 21, Appl
31	46	23.8	591	3	US-09-134-852-21	Sequence 21, Appl
32	46	23.8	650	1	US-08-325-071-56	Sequence 56, Appl
33	46	23.8	650	1	US-08-325-071-59	Sequence 59, Appl
34	46	23.8	650	4	US-08-461-004A-56	Sequence 56, Appl
35	46	23.8	650	4	US-08-461-004A-59	Sequence 59, Appl
36	46	23.8	688	1	US-08-325-071-57	Sequence 57, Appl
37	46	23.8	688	4	US-08-461-004A-57	Sequence 57, Appl
38	45.5	23.6	838	4	US-09-235-451-2	Sequence 2, Appli
39	45	23.3	82	1	US-08-319-554A-6	Sequence 6, Appli
40	45	23.3	82	1	US-08-624-123-6	Sequence 6, Appli
41	45	23.3	82	2	US-09-006-156-6	Sequence 6, Appli
42	45	23.3	147	1	US-08-325-071-69	Sequence 69, Appl
43	45	23.3	147	4	US-08-461-004A-69	Sequence 69, Appl
44	44.5	23.1	200	4	US-09-068-740A-2	Sequence 2, Appli
45	44.5	23.1	520	4	US-09-068-740A-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-586-719-4
; Sequence 4, Application US/09586719
; Patent No. 6294345
; GENERAL INFORMATION:
; APPLICANT: Zheng, Feng
; APPLICANT: Levin, Joshua Z.
; APPLICANT: Bauer, Michael W.
; TITLE OF INVENTION: Herbicide Target Genes and Methods
; FILE REFERENCE: PB/5-30909A
; CURRENT APPLICATION NUMBER: US/09/586,719
; CURRENT FILING DATE: 2000-06-05
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-586-719-4

Query Match 30.1%; Score 58; DB 4; Length 464;
Best Local Similarity 48.6%; Pred. No. 4.4;
Matches 17; Conservative 2; Mismatches 10; Indels 6; Gaps 3;

QY 2 EKLFKEVKLELENQKSGGLGKAQCAALWLQCASG 36
||| ||| :| :||| | ||| |
Db 296 EKLAKEVYF--LQNSEGSLAPFDC---WL-CLRG 324

RESULT 2
US-09-197-636-2
; Sequence 2, Application US/09197636
; Patent No. 6239267
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID
; APPLICANT: HAYES, PHILIP
; APPLICANT: MEADOWS, HELEN
; APPLICANT: DAVIS, JOHN
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: US
; ZIP: 19482-0980
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible

```
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/197,636
; FILING DATE: 23-NOV-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9805137.8
; FILING DATE: 12-MAR-1998
; APPLICATION NUMBER: UK 9815791.0
; FILING DATE: 21-JUL-1998
; APPLICATION NUMBER: UK 9819278.4
; FILING DATE: 03-SEP-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GP-30075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 601-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 839 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-197-636-2
```

```
Query Match 28.2%; Score 54.5; DB 4; Length 839;
Best Local Similarity 43.2%; Pred. No. 24;
Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;
```

```
QY 3 KLFKEVKLELENQKG-----SGLGKAQCAALWLQ 32
|| :||||| :|| :| || | ||
Db 319 KLHPTLKLEELTNKKGMTPLAALAAAGTGKIGVLAYILQ 355
```

RESULT 3

```
US-09-197-636-4
; Sequence 4, Application US/09197636
; Patent No. 6239267
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID
; APPLICANT: HAYES, PHILIP
; APPLICANT: MEADOWS, HELEN
; APPLICANT: DAVIS, JOHN
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: US
; ZIP: 19482-0980
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/197,636
; FILING DATE: 23-NOV-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9805137.8
; FILING DATE: 12-MAR-1998
; APPLICATION NUMBER: UK 9815791.0
; FILING DATE: 21-JUL-1998
; APPLICATION NUMBER: UK 9819278.4
; FILING DATE: 03-SEP-1998
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GP-30075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 601-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 839 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-197-636-4
```

```
Query Match 28.2%; Score 54.5; DB 4; Length 839;
Best Local Similarity 43.2%; Pred. No. 24;
Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;
```

```
QY 3 KLFKEVKLELENQKG-----SGLGKAQCAALWLQ 32
|| :||||| :|| :| || | ||
Db 319 KLHPTLKLEELTNKKGMTPLAALAAAGTGKIGVLAYILQ 355
```

RESULT 4

```
US-09-197-636-8
; Sequence 8, Application US/09197636
; Patent No. 6239267
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID
; APPLICANT: HAYES, PHILIP
; APPLICANT: MEADOWS, HELEN
; APPLICANT: DAVIS, JOHN
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: US
; ZIP: 19482-0980
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/197,636
; FILING DATE: 23-NOV-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9805137.8
; FILING DATE: 12-MAR-1998
; APPLICATION NUMBER: UK 9815791.0
; FILING DATE: 21-JUL-1998
; APPLICATION NUMBER: UK 9819278.4
; FILING DATE: 03-SEP-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GP-30075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 601-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 839 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-197-636-8

Query Match      28.2%; Score 54.5; DB 4; Length 839;
Best Local Similarity 43.2%; Pred. No. 24;
Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY 3 KLFKEVKLELENQK-----SGLGKAQCAALWLQ 32
   || :||||| :||| :||| :|||
Db 319 KLHPTLKLEELTNKKGMTPLALAAGTGKIGVLAYILQ 355

RESULT 5
US-09-235-451-34
; Sequence 34, Application US/09235451
; GENERAL INFORMATION:
; APPLICANT: Julius, David J.
; APPLICANT: Caterina, Michael J.
; APPLICANT: Brake, Anthony J.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING
; TITLE OF INVENTION: CAPSAICIN RECEPTOR AND CAPSAICIN RECEPTOR-RELATED
; TITLE OF INVENTION: POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 9076/084CIP
; CURRENT APPLICATION NUMBER: US/09/235,451
; CURRENT FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/072,151
; PRIOR FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: 08/915,461
; PRIOR FILING DATE: 1997-08-20
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 839
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-235-451-34

Query Match      28.2%; Score 54.5; DB 4; Length 839;
Best Local Similarity 43.2%; Pred. No. 24;
Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY 3 KLFKEVKLELENQK-----SGLGKAQCAALWLQ 32
   || :||||| :||| :||| :|||
Db 319 KLHPTLKLEELTNKKGMTPLALAAGTGKIGVLAYILQ 355

RESULT 6
US-08-325-071-61
; Sequence 61, Application US/08325071
; Patent No. 5587311
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York
; APPLICANT: WILLADSEN, Peter
; APPLICANT: KEMP, David Harold
; APPLICANT: SRISKANTHA, Alagacone
; APPLICANT: RIDING, George Alfred
; APPLICANT: RAND, Keith No. 5587311man
; TITLE OF INVENTION: DNA Encoding A Cell Membrane
; TITLE OF INVENTION: Glycoprotein Of A Tick Gut
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,071
; FILING DATE: 14-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,109
; FILING DATE: 17-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,368
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/242,196
; FILING DATE: 06-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU87/00401
; FILING DATE: 27-NOV-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PI4912
; FILING DATE: 16-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PI2570
; FILING DATE: 19-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PH9196
; FILING DATE: 27-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 60042/111 BIAU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 549 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-325-071-61

Query Match      26.4%; Score 51; DB 1; Length 549;
Best Local Similarity 35.7%; Pred. No. 46;
Matches 10; Conservative 6; Mismatches 8; Indels 4; Gaps 1;

QY 13 LENQKSGSLGKAQCAAL---WLQCSG 36
   |::| :| :| :| :| :| :|
Db 361 LKNQEAAYKGQNKCVKVDNLFWFQCADG 388

RESULT 7
US-08-461-004A-61
; Sequence 61, Application US/08461004A
; Patent No. 6235283
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York
; APPLICANT: WILLADSEN, Peter
; APPLICANT: KEMP, David Harold
; APPLICANT: SRISKANTHA, Alagacone
; APPLICANT: RIDING, George Alfred
; APPLICANT: RAND, Keith No. 6235283man
; TITLE OF INVENTION: DNA Encoding A Cell Membrane
; TITLE OF INVENTION: Glycoprotein Of A Tick Gut
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
```



```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,004A
; FILING DATE: 04-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/325,071
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,109
; FILING DATE: 17-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,368
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/242,196
; FILING DATE: 06-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU87/00401
; FILING DATE: 27-NOV-1987
; APPLICATION NUMBER: AU PI4912
; FILING DATE: 16-OCT-1987
; APPLICATION NUMBER: AU PI2570
; FILING DATE: 19-JUN-1987
; APPLICATION NUMBER: AU PH9196
; FILING DATE: 27-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 60042/152
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 549 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-461-004A-61

Query Match 26.4%; Score 51; DB 4; Length 549;
Best Local Similarity 35.7%; Pred. No. 46;
Matches 10; Conservative 6; Mismatches 8; Indels 4; Gaps 1;

QY 13 LENQKSGGLGKAQCAAL---WLQCASG 36
   |::: |::: |::: |::: |
Db 361 LKNQEAAYKGQKCVKVDNLFWFQCADG 388

RESULT 8
US-08-325-071-65
; Sequence 65, Application US/08325071
; Patent No. 5587311
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York
; APPLICANT: WILLADSEN, Peter
; APPLICANT: KEMP, David Harold
; APPLICANT: SRISKANTHA, Alagacone
; APPLICANT: RIDING, George Alfred
; APPLICANT: RAND, Keith No. 5587311man
; TITLE OF INVENTION: DNA Encoding A Cell Membrane
; TITLE OF INVENTION: Glycoprotein Of A Tick Gut
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,071
; FILING DATE: 14-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,109
; FILING DATE: 17-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,368
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/242,196
; FILING DATE: 06-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU87/00401
; FILING DATE: 27-NOV-1987
; APPLICATION NUMBER: AU PI4912
; FILING DATE: 16-OCT-1987
; APPLICATION NUMBER: AU PI2570
; FILING DATE: 19-JUN-1987
; APPLICATION NUMBER: AU PH9196
; FILING DATE: 27-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 60042/111 BIAU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 620 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-325-071-65

Query Match 26.4%; Score 51; DB 1; Length 620;
Best Local Similarity 35.7%; Pred. No. 52;
Matches 10; Conservative 6; Mismatches 8; Indels 4; Gaps 1;

QY 13 LENQKSGGLGKAQCAAL---WLQCASG 36
   |::: |::: |::: |::: |
Db 412 LKNQEAAYKGQKCVKVDNLFWFQCADG 439

RESULT 9
US-08-461-004A-65
; Sequence 65, Application US/08461004A
; Patent No. 6235283
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York
; APPLICANT: WILLADSEN, Peter
; APPLICANT: KEMP, David Harold
; APPLICANT: SRISKANTHA, Alagacone
; APPLICANT: RIDING, George Alfred
; APPLICANT: RAND, Keith No. 6235283man
; TITLE OF INVENTION: DNA Encoding A Cell Membrane
```

RESULT 10
US-08-325-071-63
; Sequence 63, Application US/08325071
; Patent No. 5587311
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York

RESULT 11
US-08-325-071-67
; Sequence 67, Application US/083250711
; Patent No. 5587311
; GENERAL INFORMATION:

RESOL 12
US-08-461-004A-63

13 LENKGSGLKAOCAAL---WLOCASG 36

US-08-470-260-3
; Sequence 3, Application US/08470260
; Patent No. 6077706
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful
; TITLE OF INVENTION: for Vaccines and Diagnostics
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,260
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,848
; FILING DATE: 21-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0316.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-470-260-3

Query Match 25.48; Score 49; DB 3; Length 1296;
Best Local Similarity 37.18; Pred. No. 2.1e+02;
Matches 13; Conservative 6; Mismatches 12; Indels 4; Gaps 2;

QY 6 KEVKLE---ELENQKSGGLG-KAQCAALWLQASG 36
| : : : | : | | | | | : |
Db 193 KNISIDNFEINNRRVGGAGRKASSTVLTQASEG 227

Search completed: October 24, 2002, 09:30:13
Job time : 6.67153 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:26:16 ; Search time 7.0219 Seconds
(without alignments)
506.317 Million cell updates/sec

Title: US-09-894-030-3_COPY_1_37
Perfect score: 193
Sequence: 1 MEKLFKEVKLELENQKSGSLGKAQCAALWLQCASGG 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues 283138
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	193	100.0	56	T12783	sublancin 168 prec
2	61	31.6	860	S43846	xylanase B - rumen
3	58	30.1	464	S61429	cystathionine beta
4	56	29.0	202	D84607	hypothetical prote
5	54.5	28.2	839	JC7621	capsaicin receptor
6	53	27.5	1000	T13636	probable minor str
7	53	27.5	1006	T13331	probable tail prot
8	52	26.9	469	AE1220	ethanolamine utili
9	52	26.9	761	T05299	hypothetical prote
10	51.5	26.7	295	D90252	conserved hypothet
11	51.5	26.7	564	HMIVF6	hemagglutinin prec
12	51.5	26.7	1024	E86331	hypothetical prote
13	51.5	26.7	2584	T24158	hypothetical prote
14	51.5	26.7	2606	T24157	hypothetical prote
15	51	26.4	361	C83350	probable transcrip
16	51	26.4	533	B96777	hypothetical prote
17	51	26.4	638	F70444	threonine--trNA li
18	51	26.4	770	T50308	probable translati
19	51	26.4	1738	C84507	hypothetical prote
20	50.5	26.2	191	S54295	GTP-binding protei
21	50.5	26.2	633	T41332	casp homolog - fis
22	50	25.9	148	H71007	hypothetical prote
23	49.5	25.6	564	HMIVF5	hemagglutinin prec
24	49.5	25.6	564	HMIVF7	hemagglutinin prec
25	49.5	25.6	564	HMIVF8	hemagglutinin prec
26	49.5	25.6	564	HMIVF9	hemagglutinin prec
27	49.5	25.6	568	A46339	hemagglutinin prec
28	49.5	25.6	754	BABOH	peptide-aspartate
29	49.5	25.6	764	C86314	hypothetical prote

30	49.5	25.6	781	2	AF2516	adenylate cyclase
31	49	25.4	243	2	A99995	hypothetical prote
32	49	25.4	243	2	F85840	hypothetical prote
33	49	25.4	259	2	A48518	probable ABC-type
34	49	25.4	393	2	G70549	probable pdhC prot
35	49	25.4	407	2	C96978	tyrosyl-tRNA synth
36	49	25.4	1288	2	E71884	vacuolating cytoto
37	49	25.4	1627	2	AE2109	two-component hybr
38	48.5	25.1	111	2	C69949	phage-related prot
39	48.5	25.1	200	2	T48130	hypothetical prote
40	48.5	25.1	534	1	S66820	heterogeneous nucl
41	48.5	25.1	560	2	PQ0041	nodulation protein
42	48.5	25.1	802	2	T37754	pmt2 methyltransfe
43	48.5	25.1	1122	2	T28130	hypothetical prote
44	48.5	25.1	1287	2	B53739	vacuolating cytoto
45	48.5	25.1	1290	2	G64630	vacuolating cytoto

ALIGNMENTS

RESULT 1
T12783
sublancin 168 precursor - Bacillus subtilis phage SPBc2
C;Species: Bacillus subtilis phage SPBc2
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 24-Sep-1999
C;Accession: T12783; H69719
R;Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Mauel, C.; Karamata, D.
submitted to the EMBL Data Library, August 1997
A;Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 pro
A;Reference number: Z17583
A;Accession: T12783
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-56 <LAZ>
A;Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025497; PIDN:AAC12992.1
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, K.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033
A;Accession: H69719
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-56 <KUN>
A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14066.1; PID:el1835
A;Experimental source: strain 168
C;Genetics: <LAL>
A;Gene: yolG
C;Genetics: <KUI>
A;Gene: sunA
C;Superfamily: unassigned lanthionine-containing peptides
C;Keywords: antibiotic; lanthionine
F;1-19/Domain: propeptide #status predicted <PRO>
F;20-56/Product: sublancin 168 #status predicted <MAT>
F;26-55/Disulfide bonds: #status experimental
F;33-48/Disulfide bonds: #status predicted
F;35/Modified site: dehydroalanine (Ser) #status experimental
F;38-41/Cross-link: (2S,3S,6R)-3-methyl-lanthionine (Thr-Cys) #status predicted

Query Match 100.0%; Score 193; DB 2; Length 56;
Best Local Similarity 100.0%; Pred. No. 3.4e-18;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKLFKEVKLELENQKSGSLGKGAQCAALWLCASGG 37
|||||
Db 1 MEKLFKEVKLELENQKSGSLGKGAQCAALWLCASGG 37

RESULT 2

S43846
xylanase B - rumen fungus (Neocallimastix patriciarum)
C;Species: Neocallimastix patriciarum
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 28-May-1999
C;Accession: S43846
R;Black, G.W.; Hazlewood, G.P.; Xue, G.P.; Orpin, C.G.; Gilbert, H.J.
Biochem. J. 299, 381-387, 1994
A;Title: Xylanase B from Neocallimastix patriciarum contains a non-catalytic 455-residue
A;Reference number: S43846; MUID:94226599
A;Accession: S43846
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-860 <BLA>
A;Cross-references: GB:S71569; GB:X76919; NID:g560648; PIDN:AAB30669.1; PID:g560649
C;Superfamily: Streptomyces endo-1,4-beta-xylanase A homology; fungal cellulose-binding
F;49-325/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SKY>
F;829-860/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 31.6%; Score 61; DB 2; Length 860;
Best Local Similarity 36.1%; Pred. No. 6.8;
Matches 13; Conservative 3; Mismatches 20; Indels 0; Gaps 0;

QY 2 EKLFKEVKLELENQKSGSLGKGAQCAALWLCASGG 37
::: |||
Db 801 QKVTVTVTVESQPTQGGMNQGGNCAAKWGCGGNG 836

RESULT 3

S61429
cystathionine beta-lyase (EC 4.4.1.8) precursor - Arabidopsis thaliana
N;Alternate names: protein F24I3.130
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 27-Apr-1996 #sequence_revision 27-Feb-1997 #text_change 19-May-2000
C;Accession: S61429; T47765
R;Ravanel, S.; Ruffet, M.L.; Douce, R.
Plant Mol. Biol. 29, 875-882, 1995
A;Title: Cloning of an Arabidopsis thaliana cDNA encoding cystathionine beta-lyase by fu
A;Reference number: S61429; MUID:96128029
A;Accession: S61429
A;Molecule type: mRNA
A;Residues: 1-464 <RAV>
A;Cross-references: EMBL:L40511; NID:g704396; PIDN:AAA99176.1; PID:g704397
R;Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;
submitted to the Protein Sequence Database, February 2000
A;Reference number: Z24475
A;Accession: T47765
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-464 <NYA>
A;Cross-references: EMBL:AL138655
A;Experimental source: cultivar Columbia; BAC clone F24I3
C;Genetics:
A;Map position: 3
A;Genome: nuclear
A;Introns: 21/3; 78/1; 116/3; 140/2; 171/1; 200/2; 237/3; 270/1; 297/2; 336/3; 372/3; 40
A;Note: F24I3.130
C;Superfamily: O-succinylhomoserine (thiol)-lyase
C;Keywords: carbon-sulfur lyase; chloroplast; methionine biosynthesis; phosphoprotein; P
F;1-70/Domain: transit peptide (chloroplast) #status predicted <TNP>
F;71-464/Product: cystathionine beta-lyase #status predicted <MAT>
F;278/Binding site: pyridoxal phosphate (lys) (covalent) #status predicted

Query Match 30.1%; Score 58; DB 2; Length 464;
Best Local Similarity 48.6%; Pred. No. 9.3;
Matches 17; Conservative 2; Mismatches 10; Indels 6; Gaps 3;

QY 2 EKLFKEVKLELENQKSGSLGKGAQCAALWLCASG 36
|||||
Db 296 EKLAKEVYF--LQNSEGSLAPFDC--WL-CLRG 324

RESULT 4

D84607
hypothetical protein At2g21970 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: D84607
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487

A;Accession: D84607
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-202 <STO>
A;Cross-references: GB:AE002093; NID:g4417289; PIDN:AAD20414.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g21970
A;Map position: 2

Query Match 29.0%; Score 56; DB 2; Length 202;
Best Local Similarity 43.3%; Pred. No. 7.8;
Matches 13; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

QY 4 LFKEVKLELENQKSGSLGKGAQCAAL--WL 31
|||:: |||
Db 130 LFKKLDVEGLSEAIAGLAAMGCAAMFAWL 159

RESULT 5

JC7621
capsaicin receptor, VR1 - human
C;Species: Homo sapiens (man)
C;Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001
C;Accession: JC7621
R;Cortright, D.N.; Crandall, M.; Sanchez, J.F.; Zou, T.; Krause, J.E.; White, G.
Biochem. Biophys. Res. Commun. 281, 1183-1189, 2001
A;Title: The tissue distribution and functional characterization of human VR1.
A;Reference number: JC7621; MUID:21139751; PMID:11243859
A;Contents: Dorsal root ganglia
A;Accession: JC7621
A;Molecule type: mRNA
A;Residues: 1-839 <COR>
A;Cross-references: GB:AF196175
C;Comment: This receptor, a transmembrane protein with many phosphorylation sites, is
its integrative activation by several noxious stimuli, and plays an important role in
C;Genetics:
A;Gene: vr1

A;Map position: 17p13
C;Keywords: transmembrane protein
F;201-233/Domain: ankyrin #status predicted <ANK1>
F;248-280/Domain: ankyrin #status predicted <ANK2>
F;333-365/Domain: ankyrin #status predicted <ANK3>
F;433-455/Domain: transmembrane #status predicted <TM1>
F;477-495/Domain: transmembrane #status predicted <TM2>
F;508-531/Domain: transmembrane #status predicted <TM3>
F;543-569/Domain: transmembrane #status predicted <TM4>
F;578-597/Domain: transmembrane #status predicted <TM5>
F;624-644/Region: pore loop #status predicted
F;656-684/Domain: transmembrane #status predicted <TM6>

Query Match 28.2%; Score 54.5; DB 2; Length 839;
Best Local Similarity 43.2%; Pred. No. 47;
Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY 3 KLFKEVKLELENQKG-----SGLGKAQCAALWLQ 32

Query Match	26.4%	Score 51;	DB 2;	Length 361;
Best Local Similarity	36.7%	Pred. No. 60;		
Matches 11:	Conservative	6;	Mismatches 13;	Indels 0;
				Gaps 0;

Search completed: October 24, 2002, 09:29:48
Job time : 11.0219 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:24:01 ; Search time 4.32117 Seconds
(without alignments)
331.536 Million cell updates/sec

Title: US-09-894-030-3_COPY_1_37
Perfect score: 193
Sequence: 1 MEKLFKEVKLELENQKSGLGKAQCAALWLQCASGG 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues 105224
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	58	30.1	464	1	METC_ARATH	P53780 arabidopsis
2	53.5	27.7	181	1	Y06E_BPT4	P13311 bacterioph
3	53.5	27.7	516	1	HXK2_DROME	Q9nft7 drosophila
4	51.5	26.7	564	1	HEMA_IACKA	P19695 influenza a
5	51	26.4	638	1	SYT_AQUAE	O67583 aquifex aeo
6	50.5	26.2	898	1	CI21_HUMAN	Q9ulv3 homo sapien
7	49.5	25.6	451	1	Z222_HUMAN	Q9uk12 homo sapien
8	49.5	25.6	563	1	HEMA_IAMAA	P26137 influenza a
9	49.5	25.6	564	1	HEMA_IADAI	P03443 influenza a
10	49.5	25.6	564	1	HEMA_IARUD	P19700 influenza a
11	49.5	25.6	564	1	HEMA_IASE2	P19701 influenza a
12	49.5	25.6	568	1	HEMA_IATKM	P26136 influenza a
13	49.5	25.6	754	1	ASPH_BOVIN	Q28056 bos taurus
14	49.5	25.4	259	1	PEB1_CAMJE	P45678 campylobact
15	49	25.4	678	1	CMC1_HUMAN	O75746 homo sapien
16	49	25.4	1288	1	VACA_HELPJ	Q9zkw5 helicobacte
17	49	25.4	1296	1	VAC1_HELPY	Q48247 helicobacte
18	48.5	25.1	111	1	YQCD_BACSU	P45939 bacillus su
19	48.5	25.1	248	1	DJ_DROME	O01352 drosophila
20	48.5	25.1	280	1	K1C3_XENLA	P05782 xenopus lae
21	48.5	25.1	534	1	NAB4_YEAST	Q99383 saccharomyc
22	48.5	25.1	560	1	NODU_AZOCA	Q07759 azorhizobiu
23	48.5	25.1	802	1	PWT2_SCHPO	O42832 schizosacch
24	48.5	25.1	1287	1	VAC2_HELPY	Q48245 helicobacte
25	48.5	25.1	1290	1	VACA_HELPY	P55981 helicobacte
26	48.5	25.1	1291	1	VAC4_HELPY	Q48258 helicobacte
27	48.5	25.1	104	1	YI64_ARCFU	O28415 archaeoglob
28	48	24.9	501	1	XYLB_LACPE	P21939 lactobacill
29	48	24.9	1043	1	RAG1_HUMAN	P15918 homo sapien
30	48	24.9	257	1	SHP_HUMAN	Q15466 homo sapien
31	47.5	24.6	293	1	YNP7_CAEEL	P34560 caenorhabdi
32	47	24.4	500	1	ERR2_HUMAN	O95718 homo sapien
33	47	24.4				

ALIGNMENTS

RESULT 1

METC_ARATH	ID	METC_ARATH	STANDARD;	PRT;	464 AA.
AC	P53780;				
DT	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Cystathionine beta-lyase, chloroplast precursor (EC 4.4.1.8) (CBL)				
DE	(Beta-cystathionase) (Cysteine lyase).				
GN	AT3G57050 OR F24I3.130.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
OX	NCBI_TaxID=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV. COLUMBIA;				
RX	MEDLINE=96128029; PubMed=8541513;				
RA	Ravanel S., Ruffet M.L., Douce R.;				
RT	"Cloning of an Arabidopsis thaliana cDNA encoding cystathionine beta-				
RT	lyase by functional complementation in Escherichia coli.";				
RL	Plant Mol. Biol. 29:875-882(1995).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV. Columbia;				
RX	MEDLINE=21016720; PubMed=11130713;				
RA	Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M., Obermaier B.,				
RA	Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Puigdomenech P.,				
RA	Deiseny M., Boutry M., Grivell L.A., Mache R., Robert C., Brottier P.,				
RA	De Simone V., Choisme N., Artiguenave F., Robert C., Brottier P.,				
RA	Wincker P., Cattolico L., Weissbach J., Saurin W., Quetier F.,				
RA	Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,				
RA	Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,				
RA	Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,				
RA	Vezi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,				
RA	Conrad A., Hornischer K., Kauer G., Loehnert T.-H., Nordsiek G.,				
RA	Reichelt J., Scharfe M., Schoen O., Bagues M., Terol J., Climent J.,				
RA	Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,				
RA	Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,				
RA	de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,				
RA	Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,				
RA	Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,				
RA	Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,				
RA	Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,				
RA	Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,				
RA	Pai G., Militischer J., Sellers P., Gill J.E., Feldblyum T.V.,				
RA	Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,				
RA	Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,				
RA	Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,				
RA	Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,				
RA	Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,				
RA	Watanabe A., Yamada M., Yasuda M., Tabata S.;				
RT	"Sequence and analysis of chromosome 3 of the plant Arabidopsis				
RT	thaliana.";				
RL	Nature 408:820-822(2000).				

P17519 potato leaf
P11622 potato leaf
P46537 bacillus ca
P18206 homo sapien
Q64727 mus musculu
P07648 escherichia
P34358 caenorhabdi
P24733 aequipecten
Q10043 caenorhabdi
Q27355 caenorhabdi
P26094 influenza a
P26095 influenza a

34 47 24.4 639 1 V70K_PLRV1
35 47 24.4 639 1 V70K_PLRVW
36 47 24.4 1065 1 CARB_BACCL
37 47 24.4 1065 1 VINC_HUMAN
38 47 24.4 1065 1 VINC_MOUSE
39 47 24.4 1122 1 EX5C_ECOLI
40 47 24.4 1691 1 YLH4_CAEEL
41 47 24.4 1938 1 MYS_AEQIR
42 46.5 24.1 325 1 YRP1_CAEEL
43 46.5 24.1 472 1 LI26_CAEEL
44 46.5 24.1 570 1 HEMA_IAHC6
45 46.5 24.1 570 1 HEMA_IAHC7

CC -!- CATALYTIC ACTIVITY: Cystathionine + H(2)O = L-homocysteine + NH(3)
CC + pyruvate.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -!- PATHWAY: THIRD STEP IN METHIONINE BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: BELONGS TO THE TRANS-SULFURATION ENZYMES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; L40511; AAA99176.1; -.
DR EMBL; AL138655; CAB72175.1; -.
DR HSSP; P00935; ICS1.
DR InterPro; IPR000277; Cys_Met_Meta_PP.
DR Pfam; PF01053; Cys_Met_Meta_PP; 1.
DR PROSITE; PS00868; CYS_MET_METAB_PP; 1.
KW Methionine biosynthesis; Lyase; Pyridoxal phosphate; Chloroplast;
KW Transit peptide.
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
FT CHAIN ? 464 CYSTATHIONINE BETA-LYASE.
FT BINDING 278 278 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 464 AA; 50429 MW; DE756848549D2CA6 CRC64;

Query Match 30.1%; Score 58; DB 1; Length 464;
Best Local Similarity 48.6%; Pred. No. 2.5;
Matches 17; Conservative 2; Mismatches 10; Indels 6; Gaps 3;

QY 2 EKLFEVKLELENQKSGGLGKAQCAALWLQCASG 36
||| ||| | : | : | | | | | | | | | | |
DB 296 EKLAKEVYF--LQNSEGSGGLAPFDC---WL-CLRG 324

RESULT 2
Y06E_BPT4
ID Y06E_BPT4 STANDARD; PRT; 181 AA.
AC P13311;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical 20.7 kDa protein in vs-regB intergenic region.
GN Y06E OR VS.1 OR 61.4.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like phages.
OX NCBI_TaxID=10665;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87066735; PubMed=3024113;
RA Valerie K., Stevens J., Lynch M., Henderson E.E., de Riel J.K.;
RT "Nucleotide sequence and analysis of the 58.3 to 65.5-kb early region
of bacteriophage T4.";
RL Nucleic Acids Res. 14:8637-8654(1986).
RN [2]
RP SEQUENCE FROM N.A.
RA Kutter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,
RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;
RA "Bacteriophage T4 genome analysis.";
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RL -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04567; CAA28228.1; -.
CC

DR EMBL; AF158101; AAD42672.1; -.
KW Hypothetical protein.
SQ SEQUENCE 181 AA; 20681 MW; 609705684A6F99B0 CRC64;

Query Match 27.7%; Score 53.5; DB 1; Length 181;
Best Local Similarity 41.7%; Pred. No. 4;
Matches 15; Conservative 5; Mismatches 15; Indels 1; Gaps 1;

QY 2 EKLFEVKLELENQKSGGLGKAQCAALWLQCASGG 37
| : | : | : | | | | | | | | | | | | | | | |
DB 39 EQFSKDGKYKTHKNHKSGLGHIMAAILW-QESSGG 73

RESULT 3
HXK2_DROME
ID HXK2_DROME STANDARD; PRT; 516 AA.
AC Q9NFT7; Q9NFT8; Q9VBF1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hexokinase type 2 (EC 2.7.1.1).
GN HEX-T2 OR HEX OR CG5443.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VARIOUS STRAINS;
RA Duvernelli D.D., Eanes W.F.;
RT "Contrasting molecular population genetics of four hexokinases in
Drosophila melanogaster and Drosophila simulans.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-K;
RA Deobagkar D.D., Kulkarni G.V., Deobagkar D.N.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
RA Fosler C., Gabrielson A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- CATALYTIC ACTIVITY: ATP + D-hexose = ADP + D-hexose 6-phosphate.
CC -!- PATHWAY: FIRST STEP OF SEVERAL METABOLIC PATHWAYS.
CC -!- SIMILARITY: BELONGS TO THE HEXOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF257590; AAG22892.1; ALT_SEQ.
DR EMBL; AF257591; AAG22894.1; ALT_SEQ.
DR EMBL; AF257592; AAG22896.1; ALT_SEQ.
DR EMBL; AF257593; AAG22898.1; ALT_SEQ.
DR EMBL; AF257594; AAG22900.1; ALT_SEQ.
DR EMBL; AF257595; AAG22902.1; ALT_SEQ.
DR EMBL; AF257596; AAG22904.1; ALT_SEQ.
DR EMBL; AF257597; AAG22906.1; ALT_SEQ.
DR EMBL; AF257598; AAG22908.1; ALT_SEQ.
DR EMBL; AF257599; AAG22910.1; ALT_SEQ.
DR EMBL; AF257600; AAG22912.1; ALT_SEQ.
DR EMBL; AF257601; AAG22914.1; ALT_SEQ.
DR EMBL; AF257602; AAG22916.1; ALT_SEQ.
DR EMBL; AF257603; AAG22918.1; ALT_SEQ.
DR EMBL; AF257604; AAG22920.1; ALT_SEQ.
DR EMBL; AF257605; AAG22922.1; ALT_SEQ.
DR EMBL; AF257606; AAG22924.1; ALT_SEQ.
DR EMBL; AF257607; AAG22926.1; ALT_SEQ.
DR EMBL; AF257608; AAG22928.1; ALT_SEQ.
DR EMBL; AJ271350; CAB67701.1; ALT_SEQ.
DR EMBL; AJ271350; CAB72132.1; ALT_SEQ.
DR EMBL; AE003756; AAF56591.1; ALT_SEQ.
DR FlyBase; FBgn0042710; Hex-t2.
DR InterPro; IPR001312; Hexokinase.
DR Pfam; PF00349; hexokinase; 1.
DR PRINTS; PR00475; HEXOKINASE.
DR ProDom; PD001109; Hexokinase; 1.
DR PROSITE; PS00378; HEXOKINASES; 1.
KW Transferase; Kinase; Glycolysis; ATP-binding;
KW Multigene family; Polymorphism.
FT BINDING 158 158 ATP.
FT DOMAIN 197 223 GLUCOSE-BINDING (POTENTIAL).
FT VARIANT 197 197 P -> A (IN STRAIN BERKELEY).
FT VARIANT 224 224 S -> N (IN STRAIN HFL97E3_15).
FT VARIANT 282 282 S -> T (IN STRAIN ZIM(S)E3_24).
FT VARIANT 410 410 S -> N (IN STRAINS SC96E3_12.3 AND
FT ZIM(S)E3_35).
FT VARIANT 514 516 SKL -> NKI (IN STRAINS DPF96E3_23.1,
FT SC96E3_12.3, HFL97E3_8, HFL97E3_12,
FT HFL97E3_16, ZIM(S)E3_24 AND ZIM(S)E3_35).
FT MISSING (IN REF. 1).
FT CONFLICT 12 17 MISSING (IN REF. 1).
SQ SEQUENCE 516 AA; 57122 MW; 01A0492BCDD567C CRC64;

Query Match 27.7%; Score 53.5; DB 1; Length 516;
Best Local Similarity 44.4%; Pred. No. 11;
Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;
QY 1 MEKLFKE-VKLEENKSGSLGKQAQCAALWLQCA 35
|:|:| | | | | | | | | | | | | | | | |
Db 479 MKKLLKPGVKFELVSEDSGRGAALVAATAVQAKS 514

RESULT 4

HEMA_IACKA STANDARD; PRT; 564 AA.
ID HEMA_IACKA
AC P19695;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Chicken/Alabama/1/75).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11338;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204912; PubMed=2705304;
RA Donis R.O., Bean W.J., Kawaoka Y., Webster R.G.;
RT "Distinct lineages of influenza virus H4 hemagglutinin genes in
RT different regions of the world.";
RL Virology 169:408-417(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M25288; AAA43221.1; -.
DR PIR; F34214; HMIVF6.
DR HSSP; P03437; 2HMG.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 342 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 344 564 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 18 18 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 564 AA; 63264 MW; B267C2A13DC1931C CRC64;
Query Match 26.7%; Score 51.5; DB 1; Length 564;
Best Local Similarity 36.4%; Pred. No. 22;
Matches 12; Conservative 7; Mismatches 11; Indels 3; Gaps 1;
QY 1 MEKLFKEVKLEENKSGSLGKQAQCAALWLQC 33
|:|:|:| | | | | | | | | | | | | | | | |
Db 458 MDKLFERVRRQLRENAEDKGN---CFEIFHC 487

RESULT 5

SYT_AQUAE STANDARD; PRT; 638 AA.
ID SYT_AQUAE
AC O67583;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Threonyl-tRNA synthetase (EC 6.1.1.3) (Threonine--trNA ligase)
DE (ThRS).
GN THRS OR AQ_1667.
OS Aquifex aeolicus.

OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -!- CATALYTIC ACTIVITY: ATP + L-threonine + tRNA(Thr) = AMP +
CC diphosphate + L-threonyl-tRNA(Thr).
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000750; AAC07549.1; -.
CC HSSP; P00955; LEVL.
CC InterPro; IPR002106; AA_trna_ligase_II.
CC InterPro; IPR004154; HGTP_anticonodon.
CC InterPro; IPR004095; TGS.
CC InterPro; IPR002314; tRNA-synt_2b.
CC InterPro; IPR002320; tRNA-synt_thr.
CC Pfam; PF03129; HGTP_anticonodon; 1.
CC Pfam; PF02824; TGS; 1.
CC Pfam; PF00587; tRNA-synt_2b; 1.
CC PRINTS; PR01047; TRNASYNTHTR.
CC PROSITE; PS00179; AA_TRNA_LIGASE_II_1; 1.
CC PROSITE; PS00339; AA_TRNA_LIGASE_II_2; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Metal-binding; Zinc; Complete proteome.
CC DOMAIN 243 536 CATALYTIC.
CC METAL 336 336 ZINC (CATALYTIC) (BY SIMILARITY).
CC METAL 387 387 ZINC (CATALYTIC) (BY SIMILARITY).
CC METAL 513 513 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 638 AA; 74097 MW; 5ECBA8227C047955 CRC64;

Query Match 26.4%; Score 51; DB 1; Length 638;
Best Local Similarity 47.8%; Pred. No. 29;
Matches 11; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 MEKLFKEVKLEENQKSGGLGK 23
Db 1 MEKIKVKIKGKEYEVEKGTPLGK 23

RESULT 6
CIZ1_HUMAN
ID CIZ1_HUMAN STANDARD; PRT; 898 AA.
AC Q9ULV3; Q9Y3G0; Q9Y3F9; Q9UHK4; Q9NYM8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cipl-interacting zinc finger protein (Nuclear protein NP94).
GN CIZ1 OR NP94 OR LSFRI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20001941; PubMed=10529385;

RA Mitsui K., Matsumoto A., Ohtsuka S., Ohtsubo M., Yoshimura A.;
RT "Cloning and characterization of a novel p21(Cip1/Waf1)-interacting
RT zinc finger protein, ciz1.";
RL Biochem. Biophys. Res. Commun. 264:457-464(1999).
RN [2]
RP SEQUENCE OF 2-898 FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Medulloblastoma;
RA Warder D.E., Keherly M.J.;
RT "NP94, a novel nuclear protein identified in a human
RT medulloblastoma.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 649-898 FROM N.A.
RX MEDLINE=99299247; PubMed=10369878;
RA Gilley J., Fried M.;
RT "Extensive gene order differences within regions of conserved synteny
RT between the Fugu and human genomes: implications for chromosomal
RT evolution and the cloning of disease genes.";
RL Hum. Mol. Genet. 8:1313-1320(1999).
CC -!- FUNCTION: MAY REGULATE THE SUBCELLULAR LOCALIZATION OF CIP/WAF1.
CC -!- SUBUNIT: INTERACTS WITH CIP/WAF1.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2/NP94B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: CONTAINS 1 MATRIN-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB030835; BAA85783.1; -.
CC EMBL; AF159025; AAF23231.1; ALT_INIT.
CC EMBL; AF234161; AAF37882.1; ALT_INIT.
CC EMBL; Y17453; CAB44346.1; -.
CC EMBL; Y17454; CAB44347.1; -.
CC InterPro; IPR003604; Znf_U1.
CC InterPro; IPR000690; Znf_matrin.
CC InterPro; IPR000822; Znf-C2H2.
CC SMART; SM00355; Znf_U1; 3.
CC SMART; SM00451; Znf_U1; 3.
CC PROSITE; PS00171; ZF_MATRIN; 1.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_2.
KW Zinc-finger; Nuclear protein; Alternative splicing.
FT DOMAIN 4 44 GLN-RICH.
FT DOMAIN 276 470 GLN-RICH.
FT ZN_FING 799 830 MATRIN-TYPE.
FT DOMAIN 741 761 GLU-RICH.
FT VARSPLIC 97 120 MISSING (IN ISOFORM NP94B).
FT VARSPLIC 197 201 MISSING (IN ISOFORM NP94B).
FT CONFLICT 9 9 MISSING (IN REF. 2).
FT CONFLICT 35 35 L -> S (IN REF. 2).
FT CONFLICT 232 232 L -> P (IN REF. 2).
FT CONFLICT 500 500 G -> S (IN REF. 2).
FT CONFLICT 555 555 G -> S (IN REF. 2).
FT CONFLICT 568 568 P -> L (IN REF. 2).
FT CONFLICT 634 634 S -> P (IN REF. 2).
FT CONFLICT 677 677 K -> R (IN REF. 2).
FT CONFLICT 698 698 R -> L (IN REF. 3).
FT CONFLICT 735 735 D -> G (IN REF. 2).
FT CONFLICT 810 810 S -> N (IN REF. 2).
SQ SEQUENCE 898 AA; 100031 MW; 28C7D76C6D4CC66B CRC64;

Query Match 26.2%; Score 50.5; DB 1; Length 898;
Best Local Similarity 35.5%; Pred. No. 46;
Matches 11; Conservative 5; Mismatches 10; Indels 5; Gaps 1;

QY 1 MEKLFKE-----VKLELENQKSGGLGKAQC 26
Db 501 MEKTLPEPVGTVSMEEIQNESACGLDVGEC 531


```
RESULT 7
2222_HUMAN
ID 2222_HUMAN STANDARD; PRT; 451 AA.
AC Q9UK12;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 222.
GN ZNF222.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shannon M., Branscomb E., Hauser L., Gordon L., Ashworth L.,
RA Stubbs L.;
RT "Differential expansion of homologous zinc-finger gene families in
RT human chromosome 19q13.2 and mouse chromosome 7.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF187988; AAF04104.1; -.
CC HSSP; P15822; 1BBO.
CC InterPro; IPR001909; KRAB.
CC InterPro; IPR000822; Znf-C2H2.
CC Pfam; PF01352; KRAB; 1.
CC Pfam; PF00096; zf-C2H2; 8.
CC SMART; SM00349; KRAB; 1.
CC SMART; SM00355; Znf_C2H2; 8.
CC PROSITE; PS50805; KRAB; 1.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
CC PROSITE; PS50157; ZINC_FINGER_C2H2_2; 10.
CC Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 8 85 KRAB.
FT DOMAIN 143 391 ZINC FINGERS.
FT ZN_FING 145 167 C2H2-TYPE.
FT ZN_FING 173 195 C2H2-TYPE.
FT ZN_FING 201 223 C2H2-TYPE.
FT ZN_FING 229 251 C2H2-TYPE.
FT ZN_FING 257 279 C2H2-TYPE.
FT ZN_FING 313 335 C2H2-TYPE.
FT ZN_FING 341 363 C2H2-TYPE.
FT ZN_FING 369 391 C2H2-TYPE.
SQ SEQUENCE 451 AA; 52532 MW; 55F55A778116D184 CRC64;

Query Match 25.6%; Score 49.5; DB 1; Length 451;
Best Local Similarity 30.6%; Pred. No. 33;
Matches 15; Conservative 4; Mismatches 13; Indels 17; Gaps 2;

QY 3 KLFKEVKLELENQKSGGLGKAQ-----CAALWLQCAS 35
||::||| | | | | | | | | | | | | | |
Db 31 KLYRDVMLENFRNLLSVG-GKIQFEMETFPPEAGTHEEFSCQIWEQIAS 78

RESULT 8
HEMA_IAMAA
ID HEMA_IAMAA STANDARD; PRT; 563 AA.
```

```
AC P26137;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Mallard/Astrakhan/244/82) (Influenza A
OS virus (strain A/Mallard/Gurjev/244/82)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11433;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91049442; PubMed=2238469;
RA Kawaoka Y., Yamnikova S., Chambers T.M., Lvov D.K., Webster R.G.;
RT "Molecular characterization of a new hemagglutinin, subtype H14, of
RT Influenza A virus.";
RL Virology 179:759-767(1990).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M35996; -; NOT_ANNOTATED_CDS.
CC HSSP; P03437; 1HTM.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT NON_TER 1 1
FT SIGNAL <1 12
FT CHAIN 13 341 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 343 563 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 16 16 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 33 33 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 237 237 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 309 309 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 496 496 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 563 AA; 62590 MW; FBD2BA976A3D5977 CRC64;

Query Match 25.6%; Score 49.5; DB 1; Length 563;
Best Local Similarity 36.4%; Pred. No. 40;
Matches 12; Conservative 6; Mismatches 12; Indels 3; Gaps 1;

QY 1 MEKLFKEVKLELENQKSGGLGKAQCAALWLQC 33
| | | | | : | | : | | | | : | |
Db 457 MNKLFERVRRLRENAEDQGNG---CFEIFHC 486

RESULT 9
HEMA_IADAL
ID HEMA_IADAL STANDARD; PRT; 564 AA.
AC P03443;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Duck/Alberta/28/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
```



```
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 342 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 344 564 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 18 18
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 564 AA; 63165 MW; 455D17B1ADD2B82B CRC64;

Query Match 25.6%; Score 49.5; DB 1; Length 564;
Best Local Similarity 36.4%; Pred. No. 40;
Matches 12; Conservative 6; Mismatches 12; Indels 3; Gaps 1;

QY 1 MEKLFKEVKLEELNKGSGGLGKAQCAALWLQC 33
   | | | | | : | | : | | | | : | |
Db 458 MNKLFERVRRLRENAEDKNG---CFEIFHC 487

RESULT 12
HEMA_IATKM
ID HEMA_IATKM STANDARD; PRT; 564 AA.
AC P19702;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Turkey/Minnesota/833/80).
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11468;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204912; PubMed=2705304;
RA Donis R.O., Bean W.J., Kawaoka Y., Webster R.G.;
RT "Distinct lineages of influenza virus H4 hemagglutinin genes in
RT different regions of the world.";
RL Virology 169:408-417(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M25290; AAA43223.1; -
CC PIR; H34214; HMIVF8.
CC HSSP; P03437; 2HMG.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 342 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 344 564 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 18 18 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
SQ SEQUENCE 564 AA; 63249 MW; A8A3D998BBE3A58F CRC64;

Query Match 25.6%; Score 49.5; DB 1; Length 564;
Best Local Similarity 36.4%; Pred. No. 40;
Matches 12; Conservative 6; Mismatches 12; Indels 3; Gaps 1;

QY 1 MEKLFKEVKLEELNKGSGGLGKAQCAALWLQC 33
   | | | | | : | | : | | | | : | |
Db 458 MNKLFERVRRLRENAEDKNG---CFEIFHC 487

RESULT 13
HEMA_IAMAB
ID HEMA_IAMAB STANDARD; PRT; 568 AA.
AC P26136;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Mallard/Astrakhan/263/82) (Influenza A
OS virus (strain A/Mallard/Gurjev/263/82)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11434;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91049442; PubMed=2238469;
RA Kawaoka Y., Yamnikova S., Chambers T.M., Lvov D.K., Webster R.G.;
RT "Molecular characterization of a new hemagglutinin, subtype H14, of
RT influenza A virus.";
RL Virology 179:759-767(1990).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M35997; -; NOT_ANNOTATED_CDS.
CC PIR; A46339; A46339.
CC HSSP; P03437; 1HTM.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 17
FT CHAIN 18 346 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 348 568 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 242 242 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 568 AA; 63149 MW; 34D2F5444BEF6BF9 CRC64;

Query Match 25.6%; Score 49.5; DB 1; Length 568;
Best Local Similarity 36.4%; Pred. No. 41;
Matches 12; Conservative 6; Mismatches 12; Indels 3; Gaps 1;

QY 1 MEKLFKEVKLEELNKGSGGLGKAQCAALWLQC 33
```


Db 462 MNKLFVRRQLRENAEDQNG---CFEIPHC 491

RESULT 14

ASPH_BOVIN

ID ASPH_BOVIN STANDARD; PRT; 754 AA.

AC Q28056;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Aspartyl/asparaginyl beta-hydroxylase (EC 1.14.11.16) (Aspartate beta-hydroxylase) (ASP beta-hydroxylase) (peptide-aspartate beta-dioxygenase).

GN ASPH.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.

OC NCBI_TaxID=9913;

OX

RN

RP SEQUENCE FROM N.A.

RC TISSUE=Liver, and Brain;

RX MEDLINE=92332546; PubMed=1378441;

RA Jia S., Vandusen W.J., Diehl R.E., Kohl N.E., Dixon R.A.F., Elliston K.O., Stern A.M., Friedman P.A.;

RA "CDNA cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase.";

RT J. Biol. Chem. 267:14322-14327(1992).

RL

RN

RP SEQUENCE OF 289-385 AND 615-641.

RC TISSUE=Liver;

RX MEDLINE=91310689; PubMed=1856229;

RA Wang Q., Vandusen W.J., Petroski C.J., Garsky V.M., Stern A.M., Friedman P.A.;

RA "Bovine liver aspartyl beta-hydroxylase. Purification and characterization.";

RT J. Biol. Chem. 266:14004-14010(1991).

RL

CC -!- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF PROTEINS.

CC -!- CATALYTIC ACTIVITY: Peptide L-aspartate + 2-oxoglutarate + O(2) = peptide 3-hydroxy-L-aspartate + succinate + CO(2).

CC -!- COFACTOR: IRON.

CC -!- SUBUNIT: MONOMER.

CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic reticulum.

CC -!- PTM: MIGHT BE PROCESSED TO THE 56 kDa (AA 289-754) OR 52 kDa (AA 311-754) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; M91213; AAA03563.1; -.

DR InterPro; IPR001440; TPR.

DR Pfam; PF00515; TPR; 2.

KW Oxidoreductase; Dioxigenase; Iron; Transmembrane; Signal-anchor;

KW Endoplasmic reticulum.

FT DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 58 78 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).

FT DOMAIN 79 754 LUMENAL (POTENTIAL).

FT DOMAIN 9 12 POLY-GLY.

FT DOMAIN 14 21 POLY-SER.

FT DOMAIN 318 328 POLY-LYS.

FT CARBOHYD 96 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 466 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 702 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 754 AA; 84998 MW; 369593A1F0B558C8 CRC64;

Query Match 25.6%; Score 49.5; DB 1; Length 754;

Best Local Similarity 38.9%; Pred. No. 53;

Matches 14; Conservative 2; Mismatches 9; Indels 11; Gaps 1;

QY 2 EKLFKEVKLELEN-----QKSGSLGKAQC 26

Db 345 EKLRKRGKIEAVNAFEELVRKYPQSPGARYGKAQC 380

RESULT 15

PEB1_CAMJE

ID PEB1_CAMJE STANDARD; PRT; 259 AA.

AC P45678;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Major cell-binding factor precursor (CBF1) (PEB1).

GN PEB1A OR C30921C.

OS Campylobacter jejuni.

OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group; Campylobacter.

OC NCBI_TaxID=197;

OX

RN

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 55026 / 81-176;

RX MEDLINE=93366784; PubMed=8360165;

RA Pei Z., Blaser M.J.;

RA "PEB1, the major cell-binding factor of Campylobacter jejuni, is a homolog of the binding component in Gram-negative nutrient transport systems.";

RT J. Biol. Chem. 268:18717-18725(1993).

RL

RN

RP SEQUENCE FROM N.A.

RC STRAIN=NCTC 11168;

RX MEDLINE=20150912; PubMed=10688204;

RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C., Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S., Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W., Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M., Whitehead S., Barrell B.G.;

RA "The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences.";

RT Nature 403:665-668(2000).

RL

RN

RP SEQUENCE OF 27-53.

RX MEDLINE=91358413; PubMed=1885571;

RA Pei Z., Ellison R.T. III, Blaser M.J.;

RT "Identification, purification, and characterization of major antigenic proteins of Campylobacter jejuni.";

RL J. Biol. Chem. 266:16363-16369(1991).

CC -!- FUNCTION: COMMON ANTIGEN AND A MAJOR CELL ADHERENCE MOLECULE. MOST PROBABLY INVOLVED, WITH PEB1C, IN A BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM FOR AN AMINO ACID. MAY BE INVOLVED IN BINDING TO INTESTINAL CELLS.

CC -!- SUBCELLULAR LOCATION: CELL SURFACE.

CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING PROTEIN FAMILY 3.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; L13662; AAA02919.1; -.

DR EMBL; AL139076; CAB73178.1; -.

DR InterPro; IPR001638; SBP_bac_3.

DR InterPro; IPR001311; SBP_glut_receptor.

DR Pfam; PF00497; SBP_bac_3; 1.

Thu Oct 24 14:40:04 2002

DR SMART; SM00062; PBPb; 1.
DR PROSITE; PS01039; SBP_BACTERIAL_3; 1.
KW Signal; Transport; Antigen; Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 259 MAJOR CELL-BINDING FACTOR.
SQ SEQUENCE 259 AA; 28177 MW; DA132ECE87BECBD5 CRC64;

Query Match 25.4%; Score 49; DB 1; Length 259;
Best Local Similarity 41.7%; Pred. No. 23;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 4 LFKEVKLEELNQKSGGLGKAQCA 27
Db 134 VLKEKKYKSLADMKGANIGVAQAA 157

Search completed: October 24, 2002, 09:28:23
Job time : 7.32117 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:24:21 ; Search time 11.8832 Seconds
(without alignments)
538.643 Million cell updates/sec

Title: US-09-894-030-3_COPY_1_37
Perfect score: 193
Sequence: 1 MEKLFKEVKELEENQKSGGLGKAQCAALWLQCASGG 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archheap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	193	100.0	56	9 O64033	O64033 bacterioph
2	193	100.0	56	16 O34781	O34781 bacillus su
3	61	31.6	860	3 Q02290	Q02290 neocallimas
4	56	29.0	170	10 Q94F34	Q94F34 arabidopsis
5	56	29.0	202	10 Q9SJ02	Q9SJ02 arabidopsis
6	56	29.0	513	5 Q9N2W3	Q9N2W3 caenorhabdi
7	54.5	28.2	453	5 Q9GN24	Q9GN24 drosophila
8	54.5	28.2	453	5 Q9GN13	Q9GN13 drosophila
9	54.5	28.2	453	5 Q9GN05	Q9GN05 drosophila
10	54.5	28.2	511	4 Q9H303	Q9H303 homo sapien
11	54.5	28.2	839	4 Q9H304	Q9H304 homo sapien
12	54.5	28.2	839	4 Q9NY22	Q9NY22 homo sapien
13	54.5	28.2	839	4 Q9NQ74	Q9NQ74 homo sapien
14	54.5	28.2	839	4 Q9H0G9	Q9H0G9 homo sapien
15	54	28.0	634	15 Q9IEZ2	Q9IEZ2 caprine art
16	53.5	27.7	453	5 Q95U08	Q95U08 drosophila

17	53	27.5	1000	9 O80181	O80181 streptococc
18	53	27.5	1006	9 O34073	O34073 streptococc
19	52.5	27.2	729	2 Q9Z487	Q9Z487 aeromonas p
20	52.5	27.2	1322	11 Q9QZP6	Q9QZP6 mus musculu
21	52	26.9	383	10 Q9LWJ2	Q9LWJ2 oryza sativ
22	52	26.9	475	10 Q9LWJ5	Q9LWJ5 oryza sativ
23	52	26.9	713	10 Q9ASW9	Q9ASW9 arabidopsis
24	52	26.9	761	10 O82634	O82634 arabidopsis
25	52	26.9	1287	2 Q93UM9	Q93UM9 helicobacte
26	52	26.9	1288	2 Q93UM8	Q93UM8 helicobacte
27	52	26.9	1289	2 Q93UM6	Q93UM6 helicobacte
28	51.5	26.7	295	17 Q97ZB8	Q97ZB8 sulfolobus
29	51.5	26.7	443	10 Q93XB2	Q93XB2 rosa hybrid
30	51.5	26.7	453	5 Q9GTU1	Q9GTU1 drosophila
31	51.5	26.7	811	10 Q9FWG8	Q9FWG8 oryza sativ
32	51.5	26.7	1024	10 Q9FXI1	Q9FXI1 arabidopsis
33	51.5	26.7	2584	5 Q9TW88	Q9TW88 caenorhabdi
34	51.5	26.7	2606	5 Q21920	Q21920 caenorhabdi
35	51	26.4	361	16 Q9I1C0	Q9I1C0 pseudomonas
36	51	26.4	368	2 Q9L4D6	Q9L4D6 xanthomonas
37	51	26.4	514	2 Q937U6	Q937U6 borrelia he
38	51	26.4	526	10 Q9LMD4	Q9LMD4 arabidopsis
39	51	26.4	533	10 Q9SSG2	Q9SSG2 arabidopsis
40	51	26.4	558	3 Q9UUA3	Q9UUA3 schizosacch
41	51	26.4	1738	10 Q9SI41	Q9SI41 arabidopsis
42	50.5	26.2	633	3 O59795	O59795 schizosacch
43	50.5	26.2	837	4 Q9BTG3	Q9BTG3 homo sapien
44	50.5	26.2	842	4 Q9H868	Q9H868 homo sapien
45	50	25.9	148	17 O59083	O59083 pyrococcus

ALIGNMENTS

RESULT 1

O64033
ID O64033 PRELIMINARY; PRT; 56 AA.
AC O64033;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE PUTATIVE LIPOPROTEIN.
GN YOLG.
OS Bacteriophage SPBC2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=66797;
RN [1]
RP SEQUENCE FROM N.A.
RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Mauel C.,
RA Karamata D.;
RT "The complete nucleotide sequence of the Bacillus subtilis SPbetac2
RT prophage.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF020713; AAC12992.1; -
KW Lipoprotein.
SQ SEQUENCE 56 AA; 5982 MW; 79EC0BF822F9F4C0 CRC64;

Query Match 100.0%; Score 193; DB 9; Length 56;
Best Local Similarity 100.0%; Pred. No. 2.4e-19;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKLFKEVKELEENQKSGGLGKAQCAALWLQCASGG 37
|||||
Db 1 MEKLFKEVKELEENQKSGGLGKAQCAALWLQCASGG 37

RESULT 2

O34781
ID O34781 PRELIMINARY; PRT; 56 AA.
AC O34781;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)

RT "The sequence of C. elegans cosmid Y94H6A.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; AC024876; AAF60896.1; -.
DR HSSP; P20393; 1A6Y.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001628; zf-C4.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00399; ZnF_C4; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW DNA-binding; Hypothetical protein; Nuclear protein; Receptor;
KW transcription regulation; Zinc-finger.
SQ SEQUENCE 513 AA; 59164 MW; C79EDF2C65E99A6E CRC64;

Query Match 29.0%; Score 56; DB 5; Length 513;
Best Local Similarity 40.6%; Pred. No. 20;
Matches 13; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

QY 5 FKEVKLELENQKSGGLGKAQCAALWLQCASG 36
|::: |||| | | | | | | | | | |
Db 311 FLKIQEEFENCPEQGQIPAPCSAEWFQELSG 342

RESULT 7

ID Q9GN24 PRELIMINARY; PRT; 453 AA.
AC Q9GN24;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HEXOKINASE-T2.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HFL97_4S, VA96_7S, GA96_2S, AND GA96_4S;
RA Duvernell D.D., Eanes W.F.;
RT "Contrasting molecular population genetics of four hexokinases in
RT Drosophila melanogaster and Drosophila simulans.";
RL Genetics 0:0-0(2000).
DR EMBL; AF257649; AAG22968.1; -.
DR EMBL; AF257642; AAG22954.1; -.
DR EMBL; AF257644; AAG22958.1; -.
DR EMBL; AF257645; AAG22960.1; -.
DR HSSP; P19367; 1HKC.
DR FlyBase; FBgn0042855; Dsim\Hex-t2.
DR InterPro; IPR001312; Hexokinase.
DR Pfam; PF00349; hexokinase; 1.
DR PRINTS; PR00475; HEXOKINASE.
DR ProDom; PD001109; Hexokinase; 1.
DR PROSITE; PS00378; HEXOKINASES; 1.
KW Kinase.

SQ SEQUENCE 453 AA; 49856 MW; E530FBF0E717B821 CRC64;

Query Match 28.2%; Score 54.5; DB 5; Length 453;
Best Local Similarity 44.4%; Pred. No. 29;
Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

QY 1 MEKLFKE-VKLELENQKSGGLGKAQCAALWLQCAS 35

Db 416 MKKLLKPGVKFELIVSEDSGRGAALVAATAVQAKS 451
|: || | | | : : : || | | | | : |

RESULT 8
Q9GN13
ID Q9GN13 PRELIMINARY; PRT; 453 AA.
AC Q9GN13;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HEXOKINASE-T2.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HFL97_2S, VA96_8S; AND HFL97_1S;
RA Duvernell D.D., Eanes W.F.;
RT "Contrasting molecular population genetics of four hexokinases in
RT Drosophila melanogaster and Drosophila simulans.";
RL Genetics 0:0-0(2000).
DR EMBL; AF257648; AAG22966.1; -.
DR EMBL; AF257643; AAG22956.1; -.
DR EMBL; AF257647; AAG22964.1; -.
DR HSSP; P19367; 1HKB.
DR FlyBase; FBgn0042855; Dsim\Hex-t2.
DR InterPro; IPR001312; Hexokinase.
DR Pfam; PF00349; hexokinase; 1.
DR PRINTS; PR00475; HEXOKINASE.
DR ProDom; PD001109; Hexokinase; 1.
DR PROSITE; PS00378; HEXOKINASES; 1.
KW Kinase.

SQ SEQUENCE 453 AA; 49843 MW; EB93A6537F1B6504 CRC64;

Query Match 28.2%; Score 54.5; DB 5; Length 453;
Best Local Similarity 44.4%; Pred. No. 29;
Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

QY 1 MEKLFKE-VKLELENQKSGGLGKAQCAALWLQCAS 35
|: || | | | : : : || | | | | : |
Db 416 MKKLLKPGVKFELIVSEDSGRGAALVAATAVQAKS 451

RESULT 9

ID Q9GN05 PRELIMINARY; PRT; 453 AA.
AC Q9GN05;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HEXOKINASE-T2.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GA96_5S, DPF96_3S, CT96_2S, CT96_5S, CT96_6S, AND VA96_5S;
RA Duvernell D.D., Eanes W.F.;
RT "Contrasting molecular population genetics of four hexokinases in
RT Drosophila melanogaster and Drosophila simulans.";
RL Genetics 0:0-0(2000).
DR EMBL; AF257646; AAG22962.1; -.
DR EMBL; AF257637; AAG22944.1; -.
DR EMBL; AF257638; AAG22946.1; -.
DR EMBL; AF257639; AAG22948.1; -.
DR EMBL; AF257640; AAG22950.1; -.
DR EMBL; AF257641; AAG22952.1; -.
DR HSSP; P19367; 1HKB.

DR FlyBase; FBgn0042855; Dsim\Hex-t2.
DR InterPro; IPR001312; Hexokinase.
DR Pfam; PF00349; hexokinase; 1.
DR PRINTS; PR00475; HEXOKINASE.
DR ProDom; PD001109; Hexokinase; 1.
DR PROSITE; PS00378; HEXOKINASES; 1.
KW Kinase.
SQ SEQUENCE 453 AA; 49869 MW; FE87E3537F1B651F CRC64;

Query Match 28.2%; Score 54.5; DB 5; Length 453;
Best Local Similarity 44.4%; Pred. No. 29;
Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

QY 1 MEKLFKE-VKLEELNQKSGSLGKAQCAALWLQAS 35
|:|:| | | | : : : | | | | | | : |
Db 416 MKLLKPGVKFELIVSDSGRGAALVAATAVQAKS 451

RESULT 10
Q9H303 ID Q9H303 PRELIMINARY; PRT; 511 AA.
AC Q9H303;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CAPSAICIN RECEPTOR VARIANT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Cortright D.N., Peck A., Zou T.;
RT "Cloning and characterization of the human capsaicin receptor, VR1.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF196176; AAG43467.1; -.
DR InterPro; IPR002110; ANK.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REP_REGION; 1.
KW ANK repeat; Receptor; Repeat.
SQ SEQUENCE 511 AA; 56977 MW; 27BAFACEBBCE0946 CRC64;

Query Match 28.2%; Score 54.5; DB 4; Length 511;
Best Local Similarity 43.2%; Pred. No. 33;
Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY 3 KLPKEVKLEELNQK-----SGLGKAQCAALWLQ 32
|| :| | | | | | | | : | | | | | | | |
Db 319 KLHPTLKLEELTNKKGMTPLAALAAAGTGKIGVLAYILQ 355

RESULT 11
Q9H304 ID Q9H304 PRELIMINARY; PRT; 839 AA.
AC Q9H304;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CAPSAICIN RECEPTOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Cortright D.N., Peck A., Zou T.;
RT "Cloning and characterization of the human capsaicin receptor, VR1.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF196175; AAG43466.1; -.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000636; Cation_chan_non_lig.
DR InterPro; IPR002111; Cat_channel_TrpL.

DR Pfam; PF00520; ion_trans; 1.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REP_REGION; 1.
KW ANK repeat; Receptor; Repeat.
SQ SEQUENCE 839 AA; 94968 MW; EA9BF7D6023EF9F4 CRC64;

Query Match 28.2%; Score 54.5; DB 4; Length 839;
Best Local Similarity 43.2%; Pred. No. 56;
Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY 3 KLFKEVKLEELNQK-----SGLGKAQCAALWLQ 32
|| :| | | | | | | | : | | | | | | | |
Db 319 KLHPTLKLEELTNKKGMTPLAALAAAGTGKIGVLAYILQ 355

RESULT 12
Q9NY22 ID Q9NY22 PRELIMINARY; PRT; 839 AA.
AC Q9NY22;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE VANILLOID RECEPTOR 1.
GN VR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=DORSAL ROOT GANGLION;
RX MEDLINE=21124568; PubMed=11226139;
RA McIntyre P., McLatchie L., Chambers A., Phillips E., Clarke M.,
RA Savidge J., Peacock M., Shah K., Winter J., Weerasekera N., Webb M.,
RA Rang H., Bevan S., James I.;
RT "Pharmacological differences between human and rat vanilloid receptor
1 (VR1).";
RL Br. J. Pharmacol. 132:1084-1094(2001).
DR EMBL; AJ272063; CAB89866.2; -.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000636; Cation_chan_non_lig.
DR InterPro; IPR002111; Cat_channel_TrpL.
DR Pfam; PF00023; ank; 3.
DR Pfam; PF00520; ion_trans; 1.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REP_REGION; 1.
KW ANK repeat; Receptor; Repeat.
SQ SEQUENCE 839 AA; 94968 MW; 773BF59D43189968 CRC64;

Query Match 28.2%; Score 54.5; DB 4; Length 839;
Best Local Similarity 43.2%; Pred. No. 56;
Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY 3 KLFKEVKLEELNQK-----SGLGKAQCAALWLQ 32
|| :| | | | | | | | : | | | | | | | |
Db 319 KLHPTLKLEELTNKKGMPLAALAAAGTGKIGVLAYILQ 355

RESULT 13
Q9NQ74 ID Q9NQ74 PRELIMINARY; PRT; 839 AA.
AC Q9NQ74;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE VANILLOID RECEPTOR 1.
GN VR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```

[1]
RP SEQUENCE FROM N.A.
RA Kelsell R.E.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBDJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504379; PubMed=11050376;
RA Hayes P., Meadows H.J., Gunthorpe M., Harries M.H., Duckworth M.D.,
RA Cairns W., Harrison D.C., Clarke C., Ellington K., Prinjha R.K.,
RA Barton A.J., Medhurst A.D., Smith G.D., Topp S., Murdock P.,
RA Sanger G.J., Terrett J., Jenkins O., Benham C.D., Randall A.D.,
RA Glover I.S., Davis J.B.;
RT "Cloning and functional expression of a human orthologue of rat
RT vanilloid receptor-1.";
RL Pain 88:205-215(2000).
DR EMBL; AJ277028; CAB95729.1; -.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000636; Cation_chan_non_lig.
DR InterPro; IPR002111; Cat_channel_TrpL.
DR Pfam; PF00520; ion_trans; 1.
DR SMART; SM00248; ANK; 3.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REP_REGION; 1.
KW ANK repeat; Receptor; Repeat.
SQ SEQUENCE 839 AA; 94924 MW; 7142FFAE43189ECC CRC64;

Query Match 28.2%; Score 54.5; DB 4; Length 839;
Best Local Similarity 43.2%; Pred. No. 56;
Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY 3 KLFKEVKLEELEENQKG-----SGLGKAQCAALWLQ 32
   || :||||| |:|| :||| | ||
Db 319 KLHPTLKLEELTNKKGMTPLAALAAAGTGKIGVLAYILQ 355

RESULT 14
Q9H0G9
ID Q9H0G9 PRELIMINARY; PRT; 839 AA.
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 94.9 KDA PROTEIN.
GN DKF2P434K0220.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansorge W., Boecker M., Bloecker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,
RA Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a Catalog of Human Genes and Proteins: Sequencing and
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs.";
RL Genome Res. 11:422-435(2001).
DR EMBL; AL136801; CAB66735.1; -.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000636; Cation_chan_non_lig.
DR InterPro; IPR002111; Cat_channel_TrpL.
DR Pfam; PF00520; ion_trans; 1.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REP_REGION; 1.
KW ANK repeat; Hypothetical protein; Repeat.
SQ SEQUENCE 839 AA; 94950 MW; EA9BF7D603AE4760 CRC64;

Query Match 28.2%; Score 54.5; DB 4; Length 839;
Best Local Similarity 43.2%; Pred. No. 56;
```

```

Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY 3 KLFKEVKLEELEENQKG-----SGLGKAQCAALWLQ 32
   || :||||| |:|| :||| | ||
Db 319 KLHPTLKLEELTNKKGMTPLAALAAAGTGKIGVLAYILQ 355

RESULT 15
Q9IEZ2
ID Q9IEZ2 PRELIMINARY; PRT; 634 AA.
AC Q9IEZ2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SURFACE GLYCOPROTEIN PRECURSOR (FRAGMENT).
GN ENV.
OS Caprine arthritis encephalitis virus (CAEV).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11660;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=021;
RX MEDLINE=20305014; PubMed=10846103;
RA Valas S., Benoit C., Baudry C., Perrin G., Mamoun R.Z.;
RT "Variability and immunogenicity of Caprine arthritis-encephalitis
RT virus surface glycoprotein.";
RL J. Virol. 74:6178-6185(2000).
DR EMBL; AJ400719; CAB95712.1; -.
KW Signal.
FT SIGNAL 1 86 POTENTIAL.
FT CHAIN 87 >634 SURFACE GLYCOPROTEIN.
FT NON_TER 634 634
SQ SEQUENCE 634 AA; 72847 MW; 75B196E23560F23E CRC64;

Query Match 28.0%; Score 54; DB 15; Length 634;
Best Local Similarity 50.0%; Pred. No. 48;
Matches 12; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKLFEVKLEELEENQKSGGLGKAQ 25
   || :||||| |:|| :||| | :|
Db 13 EKNWKEVIIIEEEEKKGGEGCTRRQ 36

Search completed: October 24, 2002, 09:29:16
Job time : 15.8832 secs
```